

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Architecture of the 90S pre-ribosome complex from *C. thermophilum*



Martin Turk
Aus
Ljubljana, Slowenien
2016

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Martin Turk betreut und von Herrn Prof. Dr. Roland Beckmann von der Fakultät für Chemie und Pharmazie vertreten.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 15. Juni 2016

Martin Turk

Dissertation eingereicht am: 21.6.2016

1. Gutachter: Professor Dr. Roland Beckmann

2. Gutachter: Professor Dr. Mario Halič

Mündliche Prüfung am 22.7.2016

Table of Content

Table of Content.....	i
Abstract.....	1
Introduction	3
1 st PART RIBOSOMES.....	3
From Information to Proteins.....	3
Ribosomes – Small and Large Subunits	4
Ribosomes – Bacteria and Eukaryotes	5
2 nd PART BIOGENESIS.....	7
Ribosomal Biogenesis	7
Compartmentalization	7
Ribosomes in Numbers.....	7
35S rDNA Organization	8
The 90S pre-Ribosome.....	9
Co-transcriptional Processing and Modifications	11
Processing	13
Modifications	16
Assembly – Temporal Binding Order	16
3 rd PART ELECTRON MICROSCOPY	20
General Introduction	20
Aims	24
Materials and methods.....	25
Sample Preparation and Contributions	25
Negative-Stain Electron Microscopy	27
Spirit cryo-EM	27
Negative-Stain (NS) IMAGIC Workflow.....	27
ISAC Workflow.....	28
Cryo-EM of the <i>S. cerevisiae</i> 90S complex.....	28
Cryo-EM of the <i>C. thermophilum</i> 90S Complex	30
Cryo-EM Model Creation.....	31

Figure Preparation	32
Results	33
Sc90S Processing.....	33
Negative-Stain (NS) – 2D Classification	34
NS – ISAC Results.....	34
NS – RELION Results.....	35
NS – IMAGIC Results	36
NS – Comparison of the 2D Class Averages	36
NS – Maskitron.....	37
NS – Initial Model Creation	37
Spirit cryo-EM	38
Spirit cryo-EM – Initial Model Generation.....	39
NS – RELION 3D Classification	39
NS – Validation of the Models Spider	40
NS – Tilt-Pair Analysis	42
Titan cryo-EM sc90S.....	44
Titan cryo-EM – 2D Classification	46
Titan cryo-EM – 29 Å Refined Map	46
Titan cryoEM – 19 Å Refined Map	47
Titan cryo-EM – RELION 3D Classification.....	50
Ct90S pre-Ribosome	51
Ct90S 3D Map at 7.3 Å Average Resolution.....	52
Segregation.....	56
UTP-A and UTP-B	59
U3 snoRNP	62
18S	65
18S – 5' Domain	65
18S – Central Domain.....	65
18S – 3' Major Domain	66
18S – 3' Minor Domain	67
18S – Summary	68
Discussion.....	70

90S and Terminal Knobs	70
Is 90S Necessary?	70
Efficiency.....	71
Energetics	72
Structure Defines Function.....	72
β -propellers and α -solenoids	75
U3 snoRNP	75
18S Folded Stages	76
Utp10	77
α -Helices.....	77
Kre33	78
Integrative Modelling.....	78
Biogenesis Highway.....	79
Stage Specific Constructs	79
The 5' ETS Beginning	79
Assembly as a Process.....	80
Biogenesis as a New Target for Antibiotics.....	80
References.....	82
<i>Curriculum Vitae</i>	90
Acknowledgments	92

Abstract

Assembly of a single ribosome is a complex task, which requires in eukaryotes the concerted action of over 200 non-ribosomal factors and a large number of snoRNAs. The pre-rRNA undergoes co-transcriptional processing, modifications and cleavage events, which separate the biogenesis pathways of the small and large ribosomal subunits at an early stage in the nucleolus. Despite the biochemical characterization of the majority of the involved factors it is not known how these components are structurally arranged in the 90S pre-ribosome, the earliest pre-ribosomal particle in the biogenesis pathway. My work showed the cryo-EM structure of the 90S pre-ribosome from *C. thermophilum* at 7.3 Å average resolution, revealing a perforated particle comprised of 19 β-propellers located mainly at the periphery and intertwined with dsRNA helices. Most of the electron density could be accounted for and assigned to the 90S sub-complexes UTP-A, UTP-B, Mpp10, Rcl1/Bms1, U3 snoRNP as well as the partially folded pre-18S rRNA in a head-down orientation, thereby providing an overall architecture of the 90S complex.

At the foot of the *ct*90S we identified the UTP-A sub-complex, which binds the first dsRNA helices of the nascent rRNA transcript at the 5' end of the 5'ETS. The growing transcript then interacts with UTP-B components and hybridizes with the 3' hinge region of the U3 snoRNP in the body part of the complex. Following the 5'ETS sequence further we encounter the pre-18S rRNA, which will ultimately form the RNA core of the small (40S) ribosomal subunit. With respect to the 90S the pre-40S is oriented upside down, pointing towards the core of the 90S. We find the 5' domain of the pre-40S to a large extent developed and resembling the mature conformation, albeit requiring a final compaction step. Part of the central domain platform is folded, but shifted with respect to its mature position relative to the 5' domain. The head is least resembling its mature state and could be partially identified on the basis of the corresponding bound r-proteins at their mature positions and Emg1, known to modify the pre-18S rRNA in that region.

The gradual development of the 18S from 5' domain first and 3' end last is similar to how ribosomes assemble in bacteria. Taken together the data provide the first structural view into the architecture of 90S pre-ribosome, and will serve as solid foundation for guiding further biochemical experiments as well as structural studies.

Introduction

In the first part the role of the ribosome is introduced in the context of the flow of genetic information followed in the second part by an overview of the ribosomal biogenesis and the 90S pre-ribosomal particle. In the last part the structural studies of the ribosome are placed in the context of technical advances, focusing mainly on electron microscopy (EM) that considerably contributed to our structural understanding of the ribosome and more recently its biogenesis.

1st PART RIBOSOMES

From Information to Proteins

Information transfer is at the very essence of life as we know it. Information has to be in a form which is storable, accessible, transferable, and at the same time allows a certain degree of change or variability we understand as evolution. The deoxyribonucleic acid (DNA) is the molecule of choice when it comes down to storing biological information for three main reasons. First, it is chemically stable over long periods of time, which makes it a safe storage medium. Second, it allows for a high level of compression i.e. it doesn't take up much space in the limited and crowded environment inside a cell. This feature is found most prominently expressed in viruses (Lavelle, 2014). In the context of DNA packaging it is pertinent to allude to another aspect over-shadowing compaction in eukaryotes, namely the organization of DNA into higher-order chromatin structure which allows for precise regulation of transcription (Hauk and Berger, 2016). Third, DNA has the intrinsic ability of being readily replicated and repaired in case of acquiring damage, provided by the redundancy of the double stranded helix structure (Crick, 1988). Storing the information in the form of DNA makes a lot of sense. However, this information is sitting idle and has to be processed to extract its content. In a rough approximation, DNA contains the blueprints for building proteins. Proteins are the molecular machines of cells, which perform various tasks on a molecular level from catalysis, signaling,

transportation, immunity, replication, to more structural roles as components of the cytoskeleton or together with sugars as part of meshes such as the peptidoglycan that surrounds most bacteria. Proteins can assemble into dynamic super-structures to perform roles in cellular motility as components of the flagella, propelling cells towards or away from a particular stimulus. In the process of building proteins the DNA is first transcribed into messenger RNA (mRNA) which is subsequently translated into a different language of amino acids, the building blocks of proteins. To this end cells evolved large macromolecular complexes called ribosomes.

Ribosomes – Small and Large Subunits

Ribosomes are the bridge between the RNA and protein worlds. Ribosomes are ribonucleoproteins consisting of ribosomal RNA (rRNA) as well as proteins distributed in two unequal parts called the small and the large ribosomal subunits. The main task of ribosomes is to decode the information carried by the mRNA molecules and to synthesize new proteins in a process known as translation (Yonath, 2009). The genetic code conveyed by the mRNAs builds upon the sequential arrangement of nucleotides, where four different kinds of nucleotides give rise to 4^3 possible triplets called codons (Gamow, 1954; Nirenberg and Matthaei, 1961; Ohtsuka et al., 1965). Three stop codons are reserved for signifying the end of transcription and usually AUG signifies the start of a transcript. The remaining codons are used to encode the 20 commonly occurring amino acids in a redundant manner, where more than one codon can code for a particular amino acid.

The small subunit harbors the decoding center whereas the large subunit is responsible for catalyzing the synthesis of nascent protein chains. The decoding is accomplished by base-pairing of a codon on the mRNA with a complementary anti-codon of a corresponding transfer RNA (tRNA) molecule which is charged with a specific amino acid (Hoagland et al., 1958; Zamecnik and Keller, 1954). The tRNA serves as an adaptor molecule binding the mRNA situated on the small ribosomal subunit and delivering amino acids to the large ribosomal subunit, where the peptide bond formation occurs in the peptidyl transferase center (PTC) (Simonović and Steitz, 2009).

Ribosomes – Bacteria and Eukaryotes

Due to their central role in protein synthesis in all living cells ribosomes had to appear very early in evolution. Ribosomes show a high degree of evolutionary conservation, the rDNA being one of the most conserved genes known. However, divergent evolution has led to differences in their composition and size. The ribosomes found in bacteria and archaea sediment at 70S and measure approximately 200 Å in size (Williamson, 2009). The small (30S) subunit of *E. coli* contains a 16S rRNA and 21 ribosomal proteins (r-proteins). The large (50S) subunit is comprised of the 5S and 23S rRNAs and 33 r-proteins. In contrast, eukaryotic 80S ribosomes are larger compared to their bacterial and archaeal homologs and measure approximately 250 by 300 Å. The small (40S) subunit in yeast contains a longer 18S rRNA and 33 r-proteins. The large (60S) subunit has three rRNA species: the 5S, 5.8S, and 25S and 46 r-proteins (Melnikov et al., 2012) (Figure 1).

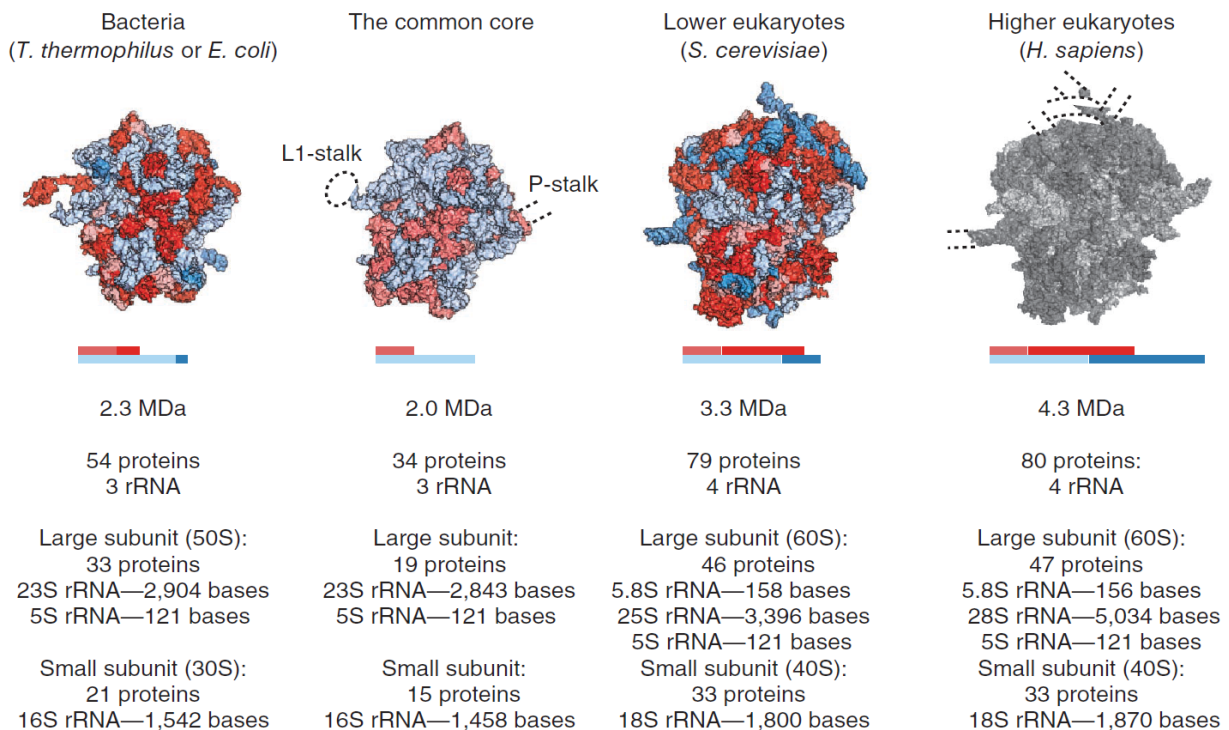


Figure 1. Ribosomes from different species and the common core they share. The common rRNA core is highlighted in light blue and proteins in light red. The expansions

to the rRNA and extensions to the proteins are colored red and blue, respectively. From Melnikov et al., 2012.

Structural studies established the view that the ribosomal core is conserved between the three domains of life and that the differences are mainly found in the surrounding shell. Eukaryotes acquired over time rRNA extension segments, new r-proteins and additional residues extending certain C- and N- termini of existing r-proteins (Woolford and Baserga, 2013). The differences between ribosomes in prokaryotes and eukaryotes are also reflected in the ratio of rRNA to proteins. In prokaryotes the ratio of rRNA to proteins is approximately 2:1, whereas in eukaryotes it is closer to one (Yonath, 2009). The main principles of how ribosomes decode the mRNA and synthesize new proteins remained conserved. However, evolution led to differences in the mechanisms of how translation is regulated. These differences are exploited by many ribosome antibiotics targeting different steps in protein synthesis (Wilson, 2014).

2nd PART BIOGENESIS

Ribosomal Biogenesis

Ribosomes are complex macromolecular machines that make proteins, but how are ribosomes made? Biogenesis of ribosomes deals with the question how ribosomes are assembled, what additional factors are involved, and how the whole process is coordinated and regulated. Biogenesis of ribosomes in all eukaryotes originates with the transcription of the rDNA resulting in a 35S pre-rRNA, a long polycistronic rRNA precursor, in the nucleolus. The 35S pre-rRNA is co-transcriptionally processed, modified and folded to yield the RNA core for the small and large ribosomal subunits. The high fidelity and efficiency of this process is accomplished by the concerted action of many protein factors and small nucleolar ribonucleoproteins (snoRNPs). These factors transiently bind to the nascent transcript and help ensure proper folding and maturation of the growing pre-ribosomal subunits as well as prevent their premature entry into the pool of translating ribosomes.

Compartmentalization

Another layer of complexity in eukaryotes is added by compartmentalization. Namely, the protein synthesis takes place in the cytosol and the transcription of the rDNA genes takes place in nucleolus inside the nucleus. All protein factors and ribosomal proteins which need to be incorporated at an early stage of ribosomal biogenesis have to be first translated in the cytosol and then transported into the nucleus and properly localized in proximity to the growing rRNA transcript in order to become successfully integrated into the growing pre-ribosomes. The pre-ribosomes have to be escorted out of the nucleus to engage in translation and the accompanying factors re-imported into the nucleus to engage in another round of ribosome biogenesis contributing to the increased number of factors involved in the process.

Ribosomes in Numbers

An exponentially growing *S. cerevisiae* produces approximately 2000 new ribosomes every minute (Warner, 1999) that have to pass through approximately 80 nuclear pore

complexes (NPC). That means 50 small or large ribosomal subunits are passing through a single NPC each minute. In order to achieve high rates of ribosome production many copies of rDNA exist. In *S. cerevisiae* there are approximately 150 such rDNA repeats located on chromosome XII (Petes, 1979). Only half of the repeats are active at any given time and are being switched on and off several times during a cell cycle (Tan et al., 2010). It was shown that active rDNA repeats are more accessible for DNA damage repair than inactive genes (Conconi et al., 2002). Employing rapid switching therefore increases the likelihood that the next generation of cells inherits intact rDNA.

35S rDNA Organization

The RNA polymerase I (Pol I) is the major polymerase dedicated to the transcription of the polycistronic 35S rRNA precursor (Kuhn et al., 2007). The 35S rDNA in *S. cerevisiae* is 6600 nucleotides long and is organized in the following way. The 18S, 5.8S and 25S rRNA are separated by internal transcribed spacers (ITS) and flanked by two external transcribed spacers (ETS). In a series of processing steps these spacers are excised for the pre-rRNAs to reach their mature state (Figure 2). The 18S will be the integral part of the small ribosomal subunit. The 5.8S and 25S will be the RNA core of the large ribosomal subunit together with the 5S rRNA, which is transcribed by the RNA polymerase III in the reverse direction on the same locus preceding the 5' ETS sequence.

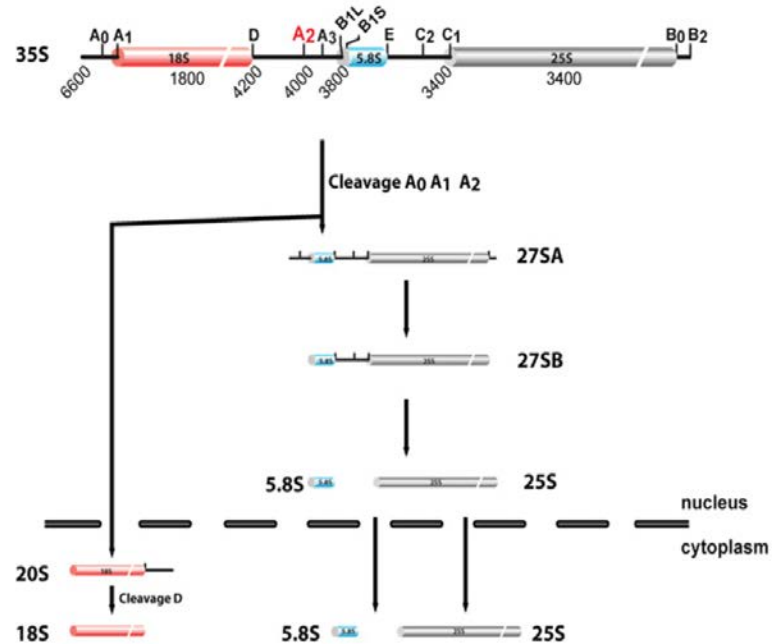


Figure 2. A simplified view of the processing steps of the 35S rRNA. The small (40S) subunit rRNA path is highlighted in red and the two rRNAs of the large (60S) subunit are shown in blue and gray. From Kos & Tollervay, 2010.

The 90S pre-Ribosome

As Pol I engages transcription of the rDNA the first to emerge is the 5' ETS sequence. The 5' end of which attracts co-transcriptionally biogenesis factors which in turn enable more factors to associate with the rapidly growing transcript before transcription of the small (40S) subunit 18S rRNA starts. As the transcript reaches the 3' end of the 18S rRNA a globular particle resembling terminal knobs in Miller spreads can be observed. This particle containing the 5' ETS, the 18S and part of the ITS1 rRNA sequences plus a number of associated proteins was first isolated in 2002 (Dragon et al., 2002) and called the small subunit (SSU) processome, the name implying its function, and almost concomitantly a very similar particle was identified called the 90S pre-ribosome, its name based on sedimentation assays (Grandi et al., 2002). The existence of such a particle dates back to the '70s (Trapman, Retel, & Planta, 1975). Biochemical characterization of the early rRNA particles was accomplished by tandem affinity pullouts linked to mass spectrometry (TAP-MS) and DNA microarrays identifying the

components of the RNA-processing complexes (Krogan et al., 2004). Following this systematic large-scale approach, classification of the overlapping components into sub-complexes became possible. Most of the approximately 70 associated 90S factors are Utp (U three proteins) that form stable sub-complexes termed UTP-A (Utp4, Utp5, Utp8, Utp9, Utp10, Utp15, Utp17), UTP-B (Utp1, Utp6, Utp12, Utp13, Utp18, Utp21), and UTP-C (Utp22, Rrp7) (Figure 3).

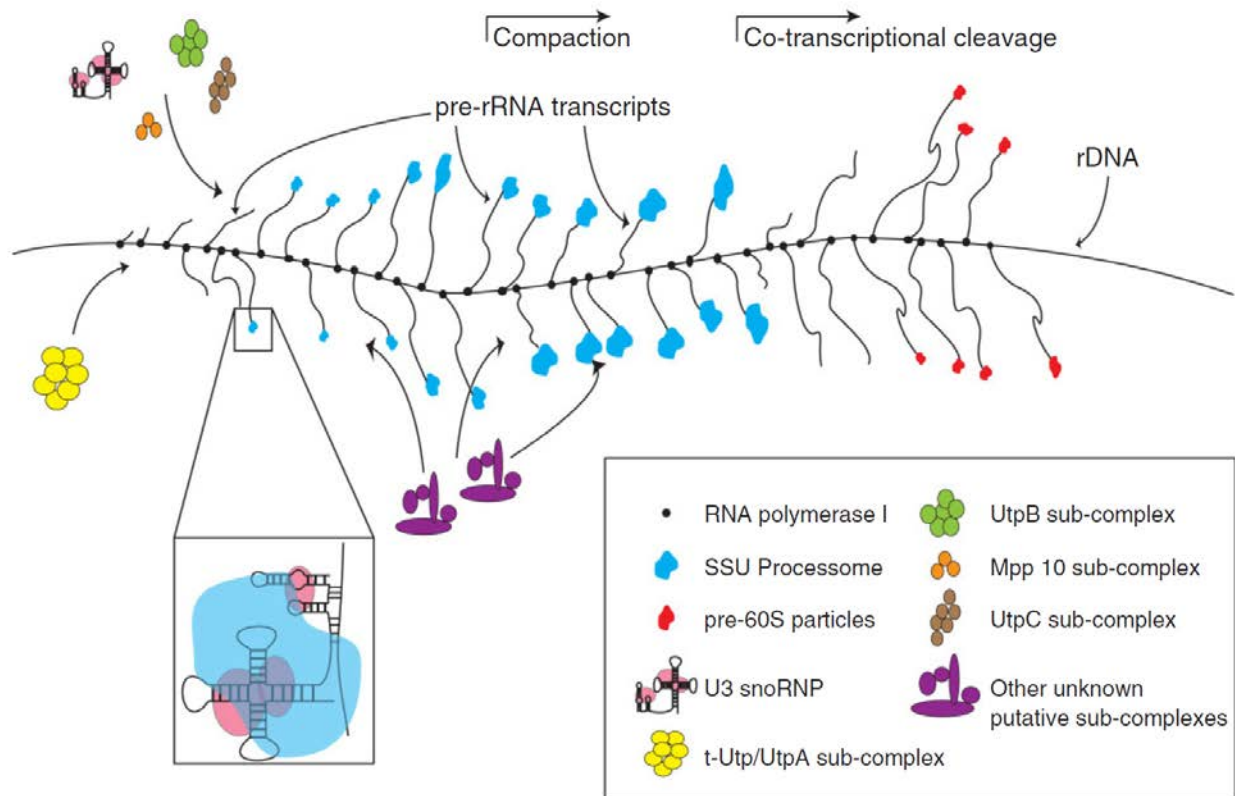


Figure 3: Schematic representation of the Miller chromatin spreads showing the sequential assembly of the 90S sub-complexes on the pre-rRNA as the transcript increases in length. From Phipps, Charette, & Baserga, 2011.

Among the 90S interacting snoRNPs only U3, U14 and snR30 are essential (Karbstein, 2011). The U3 snoRNP is recruited early on to the 5' ETS sequence forming base-pairs also at two regions in the 18S rRNA prohibiting the formation of the central pseudoknot, a key feature required for translation initiation (Brink et al., 1993) (Figure 4). The interaction between U3 and 18S (Dutca et al., 2011) is stabilized by the incorporation of the Mpp10/Imp3/Imp4 sub-module (Gérczei et al., 2009). In addition, the U3 snoRNA

base pairing to rRNA directs the early cleavages A_0 , A_1 , A_2 . It was shown the GTPase Bms1 binds to the U3 snoRNP and may act as a molecular switch, since Bms1 binds the endonuclease Rcl1 (Horn et al., 2011) responsible for the A_2 cleavage. Interestingly, the GTP hydrolysis of Bms1 is inhibited by the binding to U3 snoRNA and Rcl1, suggesting a trigger is required for initiating GTP hydrolysis (Karbstein and Doudna, 2006).

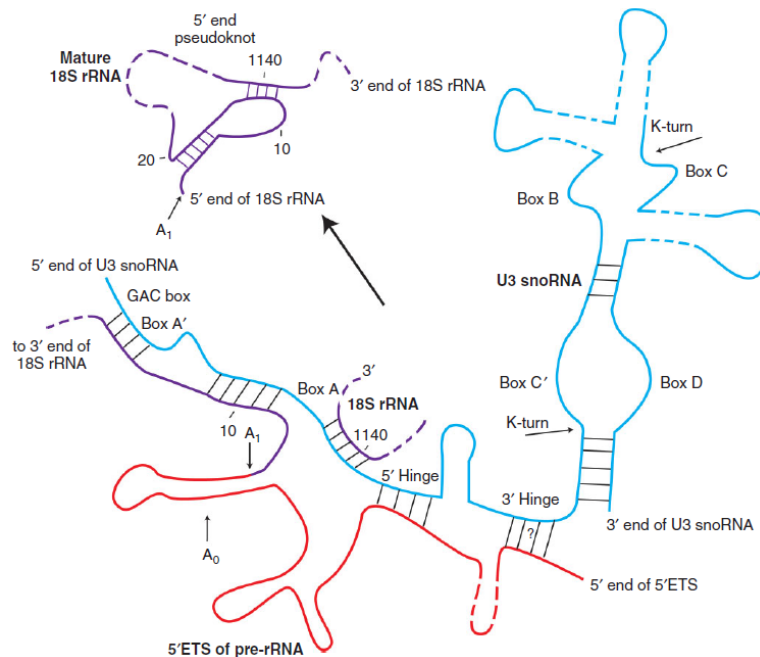


Figure 4: The binding sites of the U3 small nucleolar RNA (snoRNA) with the primary rRNA transcript. The U3 snoRNA (blue) hybridizes with the pre-18S rRNA (purple) at three sites and with the 5' ETS (red) at two sites. In addition to guiding the early cleavages A_0 , A_1 , and A_2 the U3 snoRNA has been proposed to hybridize with the pre-18S rRNA obstructing the central pseudoknot formation characteristic of the mature 18S rRNA. From Phipps et al., 2011.

Co-transcriptional Processing and Modifications

For a long time the largely accepted view suggested that the full 35S pre-rRNA is fully transcribed before any processing takes place. However, the endogenous purification and MS of the 90S complex showed there are very few 60S associated factors present in the 90S pullouts (Grandi et al., 2002) challenging this model. More recent EM

analysis of the Miller spreads in different yeast strains showed that compaction of the pre-18S rRNA into a large 90S particle, its maturation and release all occur during transcription (Osheim et al., 2004). The time-line of rRNA transcription was estimated by the 60 nucleotides/s elongation rate of Pol I in exponentially growing yeast (French et al., 2003). From the 5' ETS to the 3' end of 18S the Pol I requires approximately 55 s coinciding with the appearance of the large SSU particle in the Miller spreads. Based on the distance and average speed of the Pol I from the known cleavage in ITS1 that releases the SSU co-transcriptionally, there is a dwell time of approximately 40 s in which the particle matures before it is released from the growing transcript and cannot be further tracked by this method (Figure 5).

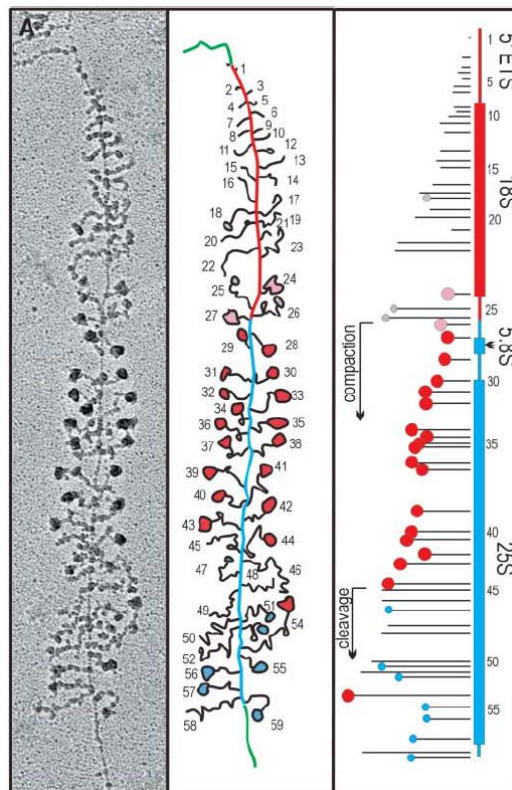


Figure 5: Transcript mapping of the Miller spreads. In red the small subunit (SSU) processome is shown and in blue the large ribosomal subunit transcripts. The particles in gray show small terminal knobs that grow into early SSU knobs in pink as the Pol I reaches the 3' end of the 18S. The mature SSU terminal knobs, which have not yet been cleaved off the primary rRNA transcript are colored red. From Osheim et al., 2004.

The development of a fast kinetic labeling method quantitatively showed that 70% of the pre-18S rRNA transcripts are cleaved co-transcriptionally i.e. before reaching the 3' end and that the modifications of the pre-rRNA also occur co-transcriptionally and not after the release of the transcripts (Kos & Tollervey, 2010). A shorter and probably more accurate transcription time of 40 nucleotides/s was estimated, but nevertheless both studies confirmed that the co-transcriptional processing and modification of the SSU particle indeed occurs before its release and that processing of the fully transcribed 35S rRNA is rather the exception.

Processing

The exact order of the first cleavages is not clear and different processing intermediates can be purified. Upon perturbing the expression levels of involved factors, different rRNA intermediates can be enriched for. Clear boundaries discriminating on- and off-pathway products are often subject of discussion. The major U3-dependent pathway results in three different sized intermediates 33S, 32S, and 20S corresponding to cleavages at positions A_0 and A_1 in the 5' ETS and cleavage A_2 in the ITS1, respectively (Figure 6). A minor pathway with the A_3 cleavage occurring first yields the 23S pre-rRNA that is subsequently processed in an U3 dependent manner to 20S pre-rRNA (Gallagher et al., 2004). The A_2 cleavage generates two fragments, one 20S precursor containing the 18S and one 27SA₂ precursor containing the 5.8S and 25S (in mammals 28S). The 20S at this stage requires one more cleavage step at the D-site to mature into the 18S. Exactly when the D-cleavage occurs is not known. The release of the ITS1 following the D-cleavage is known to occur in the cytosol and is thought to be coupled with the reorganization of the decoding site located on h44 of the 18S 3' minor domain. ITS1 is hybridized throughout the maturation to the decoding site in an alternate conformation prohibiting a pre-mature involvement of the pre-40S in translation (Lamanna and Karbstein, 2011). In contrast to the large set of factors associated with the early pre-40S, the cytosolic pre-40S contains only seven factors none of which are present in the early pre-40S particles. These late factors are inhibiting translation initiation, most notably Tsr1 by prohibiting joining of the 60S subunit (Strunk et al., 2012). Since cytosolic pre-40S were also shown to be present in polysomes (Soudet et al.,

2010) contradicting the previous observation that more than one intermediate of the cytosolic pre-40S likely exists (Karbstein, 2011).

The 27S processing of the large (60S) subunit rRNAs in the nucleus is more complex. Majority of the 27SA₂ is cleaved by mitochondrial RNA processing (MRP) RNase at position A₃ to yield the 27SA₃ pre-rRNA. This pre-RNA is subsequently shortened at the 5' end by the 5'→3' exonucleases Rat1 and Rrp17 to the 27SB_S. Only 10-15 % of the 27SA₂ pre-rRNA are cleaved by a yet unidentified endonuclease resulting in 27SB_L. Products of both parallel pathways undergo the cleavage at site C₂, which separates the two 27S rRNAs into 25.5S and 7S_S or 7S_L intermediates. The 25.5S is trimmed at the 5' end resulting in the 25S rRNA. The 7S_S and 7S_L are trimmed at the 3'end resulting in 5.8S_S and 5.8S_L rRNAs which differ only by 6 nucleotides at their 5'end. The significance of this difference is yet to be unraveled, it is however a conserved feature in eukaryotes.

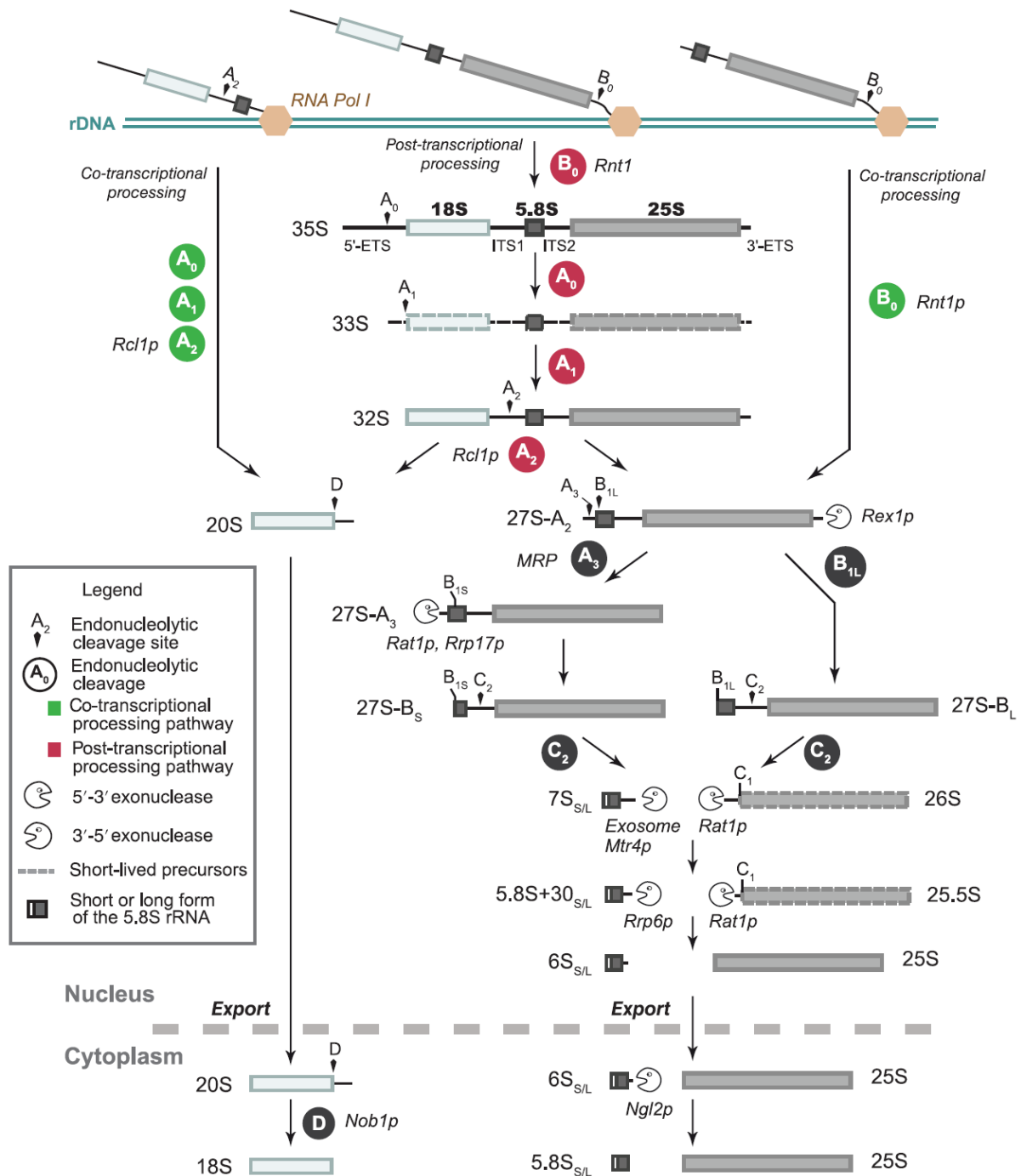


Figure 6. The processing steps of the rRNA primary transcript. The co-transcriptional early cleavages (green) result in a pre-20S rRNA, which is cleaved at the D-site by Nob1 in the cytosol resulting in the 18S rRNA, and on the other end 27SA₂. The 27SA₂ undergoes several processing steps to finally reaching a short and a long 5.8S and the 25S rRNA that will be part of the large ribosomal subunit. From Henras, Plisson-Chastang, O'Donohue, Chakraborty, & Gleizes, 2015.

Modifications

The rRNA is in addition to co-transcriptional folding and cleavages also modified at certain nucleotides. The two rRNA modifications are the pseudo-uridylation (Ψ) and 2'-hydroxyl methylation of the ribose. The modifications are carried out by snoRNPs, which are comprised of a particular snoRNA and a set of specific proteins. The role of the associated proteins is structural as well as enzymatic as they carry out the catalysis of the modifications. A particular snoRNA hybridizes with the target RNA being modified and directs or guides the site of the modifications by correctly positioning the proteins to catalyze the modification. There are two major classes of snoRNPs, the box C/D snoRNPs responsible for methylation and the box H/ACA snoRNPs performing modifications on uridines.

The box C/D snoRNPs contain conserved box C, C' and D, D' motifs and a complementary antisense sequence of 10-21 nucleotides that binds to the target RNA. The methylation is directed at the 5th nucleotide upstream of the box D or D'. The proteins which steadily associate with the box C/D snoRNA are the methyl transferase Nop1 (fibrillarin), Nop58, Nop56 that form coiled coils for oligomerization, and Snu13 a kink-turn binding C/D snoRNA. These protein components are symmetrically arranged in the structure. The U3 snoRNP is a special case of C/D box snoRNPs, because it is not involved in the transfer of a methyl group to ribose, but rather directs the early cleavages and performs a structural role in the 90S complex. In addition to the normal set of proteins it contains a WD-40 repeat containing protein Rrp9 that is helping localize the U3 snoRNP to the nucleolus.

Assembly – Temporal Binding Order

The order of assembly of the r-proteins, protein factors, the snoRNPs and the rRNA into the 90S is riddled by the complexity of the process and till recently by the lack of tools to simplify the assembly stages into discrete steps. It was shown that the hierarchy of assembly starts with the UTP-A sub-complex, which is necessary for the association of the following set of sub-complexes U3 snoRNP, UTP-B and Mpp10. Independent of the first set, UTP-A enables the incorporation of Rrp5 which is required for binding of the

sub-complex UTP-C to the 90S (Pérez-Fernández et al., 2007). The binding sites of several 90S proteins to the rRNA were mapped using UV cross-linking and analysis of cDNA (CRAC) (Granneman et al., 2009, 2010), but the temporal order of these binding events is not well understood.

The timeline of binding events of the r-proteins and rRNA folding in eukaryotic ribosome assembly has been derived to a large extent from our understanding of the biogenesis in bacteria. Bacteria have considerably less factors involved in the biogenesis of ribosomes most of which are not essential offering a simplified system (Stokes et al., 2014). Successful *in vitro* reconstitutions of the bacterial ribosomes (Held et al., 1973; Nierhaus and Dohme, 1974; Traub and Nomura, 1968) have not only shown that this is possible, but enabled a view into the sequential order of the binding events. The Nomura map shows the temporal assembly order of the 30S r-proteins in bacteria (Mizushima and Nomura, 1970). Pulse-chase labeling coupled to quantitative MS after trypsin digestion showed improved resolution of the r-protein binding rates for the 30S ribosomal subunit (Bunner et al., 2008) (Figure 7).

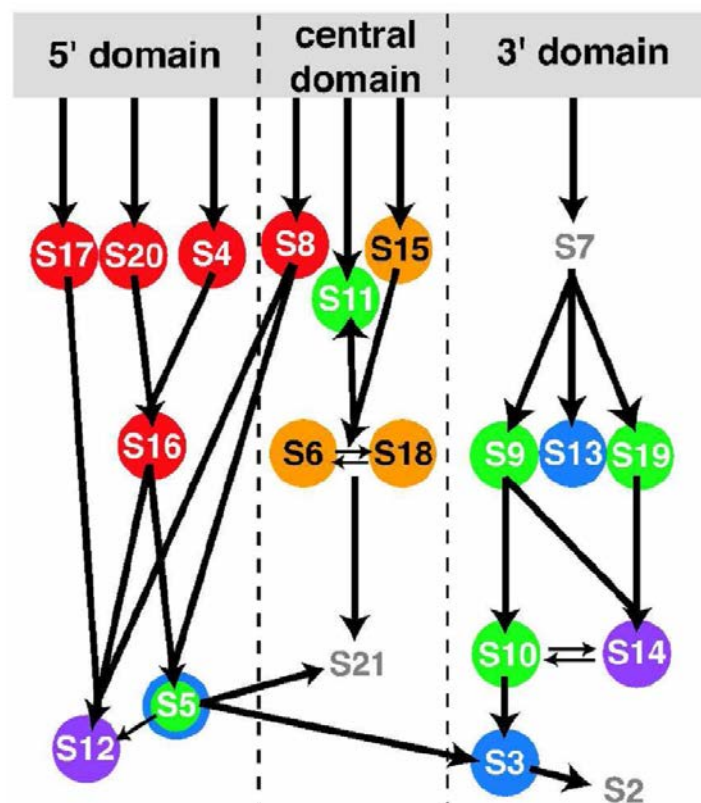


Figure 7. The Nomura assembly map. The binding rates of the pulse-chase and quantitative mass spectrometry of the 30S ribosomal subunit are indicated by colors of the circles: 20 to 30 min⁻¹ (red), 8.1 to 15 min⁻¹ (orange), 1.2 – 2.2 min⁻¹ (green), 0.38 – 0.73 min⁻¹ (blue), 0.18 – 0.26 min⁻¹ (purple). Binding rates for S2 and S21 could not be determined and are therefore shown in gray. From Bunner et al., 2008.

The elongation of the growing transcript from 5' to 3' end coincides with the order in which the r-proteins associate with the nascent transcript. The r-proteins binding to the body form a stable intermediate that subsequently acquires the platform and the head. Multiple parallel pathways were observed in the 30S ribosome assembly, not all leading to a functional ribosome (Mulder et al., 2010). The role of r-proteins in the rRNA maturation is believed to stabilize the RNA conformations which lead to productive outcomes. Since the binding of r-proteins also has an effect of the rRNA folding, the binding of early r-proteins is thought to induce conformational changes in the rRNA thus enabling the binding of later r-proteins.

It was shown the 35S rRNA transcribed from a plasmid using the RNA polymerase II can readily substitute for the loss of Pol I transcription in *S. cerevisiae* (Nogi et al., 1991). Based on this research only recently affinity tagged constructs of different length rRNA coupled to MS analysis led to a discrete method for elucidating the sequential assembly of factors involved in the early steps of the yeast small subunit formation (Figure 8). The results show in a step-wise manner which factors are recruited to sequences in the 5' ETS, the four 18S rRNA domains, and ITS1 (Chaker-Margot et al., 2015). Since there is a ribosomal contamination in the pullouts the r-proteins binding cannot be estimated and only a view of the assembly factors is provided. Interestingly, most of the 90S sub-complexes already bind to the 5' ETS (UTP-A, UTP-B, Mpp10, U3 snoRNP). Association of two proteins Utp3 and Lcp5 containing exosome interacting domains with the 5'ETS and 18S 5'domain, respectively, strategically positions both proteins at the rRNA interface for subsequent nuclear exosome driven degradation and recycling of the 5' ETS particle after the pre-40S release from the 90S particle has occurred.

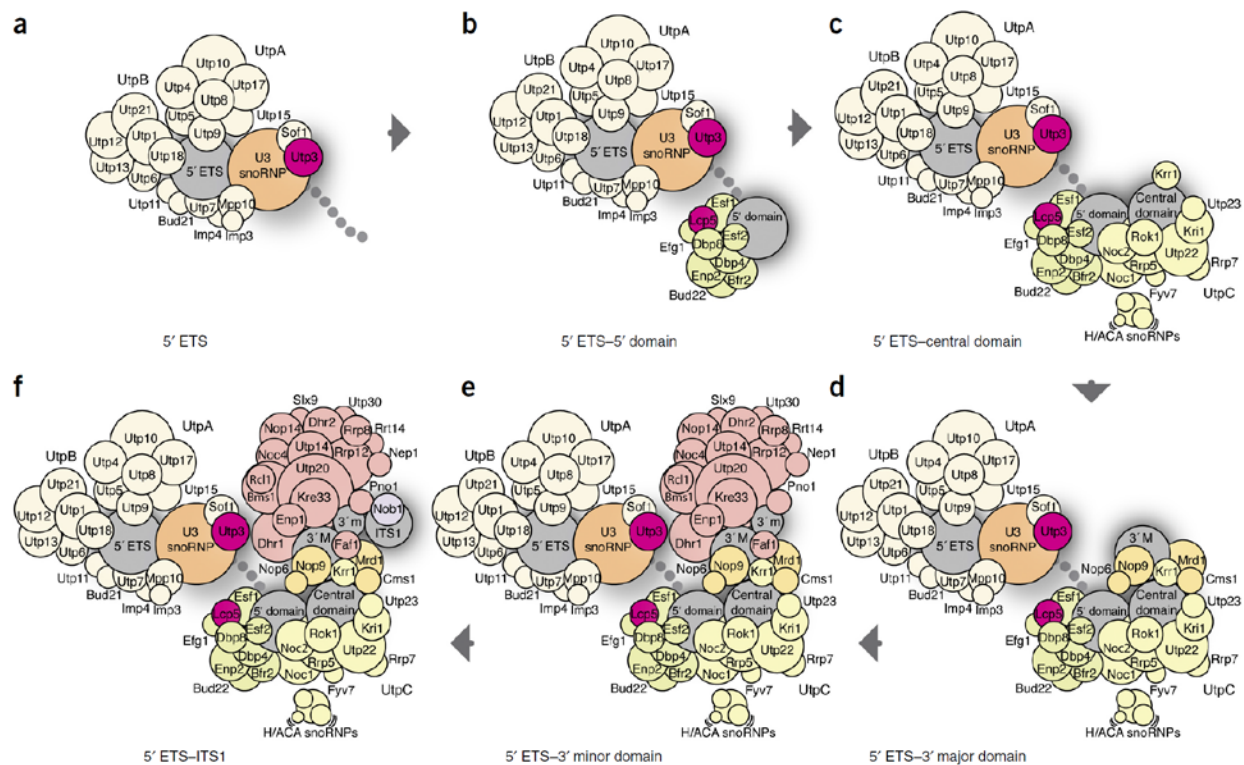


Figure 8. Different assembly stages of the 90S pre-ribosome factors. Assembly factors that were found to associate with a particular rRNA sequence are colored as follows: 5'ETS (beige), 5' domain (light green), central domain (light yellow), 3' major domain (light orange), 3' minor domain (light pink), and ITS1 (light purple). From Chaker-Margot et al., 2015.

3rd PART ELECTRON MICROSCOPY

General Introduction

Since the first electron microscope prototype was built in 1931 by Ernst Ruska and Max Knoll, electron microscopes quickly surpassed the resolution of light microscopes. It did not take long before the first commercial Siemens microscopes appeared that scientists started discovering structures inside cells that were previously not detected by light microscopy. A prominent example was the discovery of ribosomes by Georg Emil Palade (Palade, 1955). The central role of ribosomes in protein biosynthesis and their large abundance made ribosomes interesting for biochemical and structural investigations. The ribosome has since been extensively studied by both X-ray crystallography and electron microscopy (EM) and played an important role in the development of single particle cryo-EM. Interestingly, the first view into the structure of a known gene was provided by EM. The structural organization of the rDNA was pioneered by Oscar Miller and Barbara Beatty in 1969. The extra-chromosomal nucleoli from amphibian oocytes were dispersed on carbon-coated grids and could be recorded on an electron microscope after negative staining. The micrographs of the so-called Miller spreads showed that tandem repeats of rDNA were narrowly spaced (Figure 9). The micrographs showed also that the rDNA repeats were being transcribed by multiple RNA polymerases at the same time, which gave them the typical “Christmas trees” appearance as the transcripts became longer in size. The observed rRNA precursor was correctly proposed to correspond to one 18S and one 28S rRNA molecule with some additional segment that was known to be subsequently degraded during the maturation of the two molecules (Miller and Beatty, 1969). The Miller spreads were not only an important step in understanding gene organization, but also provided the first view into the birth of ribosomes.

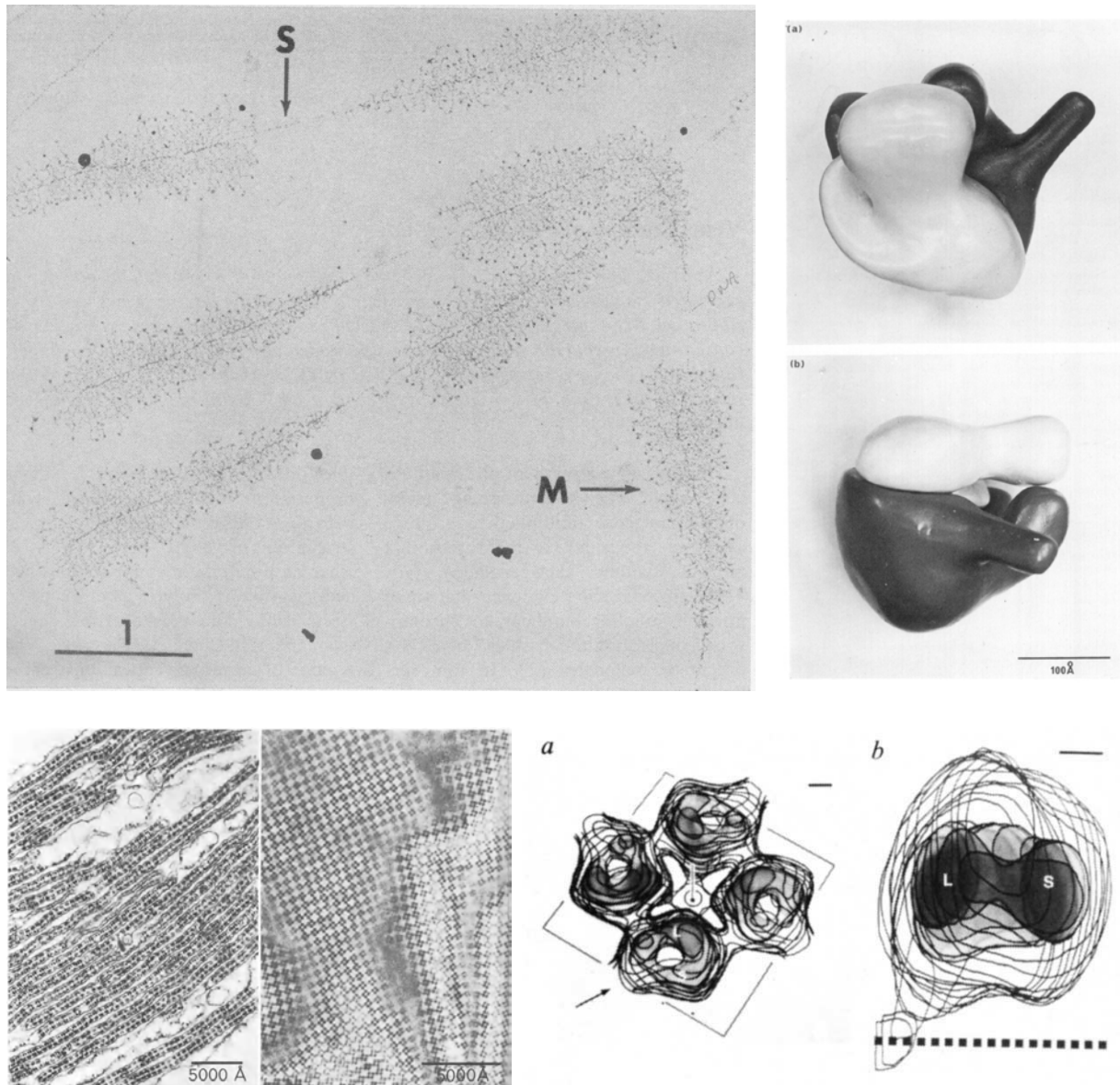


Figure 9. Top left: The Miller spreads showing matrix (M) segments and matrix-free (S) segments. From Miller & Beatty, 1969. Top right: Plasticine models conceptually grasping the small and large ribosomal subunits. From Lake, 1976. Bottom left: vertical and lateral sections of ribosome crystalline sheets. From Kühlbrandt & Unwin, 1980. Bottom right: 3D model of the membrane-bound ribosomes by electron diffraction. From Unwin, 1977.

The first 3D model of the *E. coli* ribosome emerged from carefully matching Plasticine models to negatively stained micrographs. This was done separately for the small (30S) and large (50S) subunits to reduce flattening artifacts (Lake, 1976). The next structure

of a ribosome was derived from electron crystallography. Ribosomes were observed to arrange in crystalline sheets spontaneously during oogenesis of the lizard *Lacerta sicula* under cold temperatures (Taddei, 1972). The membrane-bound ribosomes appeared to be arranged in tetramers and EM diffraction of the 2D crystal arrays was possible resulting in a low-resolution model of the ribosome (Figure 9) (Unwin, 1977). Spontaneous 2D crystal formation of ribosomes also encouraged crystallographers that obtaining 3D crystals of ribosomes is possible. First crystals of the 50S diffracting to 3.5 Å were obtained already in 1980 (A. Yonath, J. Muessig, B. Tesche, S. Lorenz, 1980). However, it took another two decades before the ribosome structure was solved to atomic resolution and the mechanism of how ribosomes function explained (Ban et al., 2000; Schlutzen et al., 2000; Wimberly et al., 2000). In addition to optimizing conditions to obtain high-quality crystals, advances in instrumentation were necessary such as the development of synchrotrons and cryo-crystallography (Hope, 1988; Hope et al., 1989). Similarly, in EM it took a number of years before the sample preparation (Dubochet et al., 1982, 1985, 1988), electron microscopes and computational methods (Carazo and Frank, 1988; Penczek et al., 1994) had been developed to a point where 3D reconstructions of non-periodic objects such as ribosomes became feasible by cryo-EM (Frank, 2016). Only then visualization of long sought complexes of high biological relevance became possible. An important example being the research on the ribosome and its interaction with the protein conducting channel (PCC) conducted at the Rockefeller Institute of Medicine over more than 40 years. The story dates back to 1955 where ribosomes were observed to be sitting on the ER membranes (Palade, 1955). The second piece followed two decades later with the postulation of the signal hypothesis suggesting the existence of ribosomes interacting with a PCC for the transfer of nascent protein chains over the ER membrane (Blobel and Dobberstein, 1975). Finally, in the late 90s, the ribosome structure in complex with the Sec61 PCC was revealed (Beckmann et al., 1997), and shortly thereafter in its active form with tRNA (Beckmann et al., 2001a, 2001b).

The biogenesis intermediates especially at the early stages of ribosome assembly are believed to be extremely dynamic and rapidly developing particles rendering isolation of highly homogenous sample difficult. In contrast, the later steps of pre-60S maturation

are well characterized where individual stages of assembly are associated with a particular factor or set of factors that interact with the growing particle in defined intervals (Kressler et al., 2010). Isolation of a pre-60S intermediate based on two overlapping factors allows narrowing down the time-window to a compositionally clearly defined assembly stage. This is not the case for the 90S pre-ribosomes where this strategy cannot be applied yet. In fact, it is believed the 90S assembly is a one-step collapse (Osheim et al., 2004). Co-transcriptional processes and transient association of factors lead to a certain degree of conformational as well as compositional heterogeneity of these growing early intermediates which makes crystallography of the 90S difficult if not impossible at this stage.

Recent advances in technology in particular the advent of direct electron detectors with an improved signal-to-noise ratio and fast readouts allowing motion correction have considerably facilitated structural determination of macromolecular complexes by EM (Li et al., 2013).

For the above reasons single particle cryo-EM was the method of choice for determining the *de novo* structure of the 90S pre-ribosomal particle in this work. The direct detector was not available for the entire *S. cerevisiae* 90S reconstruction and was used only with the *C. thermophilum* 90S pre-ribosome. Despite the sub-optimal implementation of the detector at the time of the *ct*90S data-collection the difference in the image quality compared to the *sc*90S particles was apparent. Most of the issues related to the low signal-to-noise ratio in the *sc*90S were not present for the structural determination of the *ct*90S pre-ribosome.

Aims

The aim of the project was the structural determination of the 90S pre-ribosome. To that end several steps were necessary. First, endogenous purification of a homogenous 90S sample from yeast was established. Several proteins were tagged and the yields estimated until an optimal bait protein was found yielding high amounts of homogenous sample. Second, an *ab initio* model from negative stained EM micrographs was obtained. Third, single particle cryo-EM of the sc90S and later ct90S samples was performed. For the projection-matching refinement the reconstructed initial model from the negative stain images was used. Fourth, the obtained electron density map was interpreted and a basic model of the 90S pre-ribosome architecture provided.

Materials and methods

Sample Preparation and Contributions

Several samples were prepared and analyzed during the course of the project as part of a fruitful collaboration between the Beckmann and Hurt laboratories. The *Saccharomyces cerevisiae* 90S (sc90S) samples were purified by Markus Kornprobst (Hurt laboratory) and the *Chaetomium thermophilum* (ct90S) samples by Dr. Nikola Kellner (Hurt laboratory).

Since this was not part of my own work I will only conceptually describe the purifications for preparing samples for electron microscopy without going into much detail. Detailed purification procedures can be found in the resulting publication Kornprobst, Turk et al., 2016.

For the *Saccharomyces cerevisiae* sc90S cryo-EM sample the best yield was obtained by a C-terminal Flag-TEV-protA (FTpA) tag on the UTP-A sub-complex protein Utp10. In addition to the sc90S complex, free UTP-A and the post 90S dissociation 5' ETS complex co-purified, requiring a sucrose gradient centrifugation step to isolate the fractions of interest. The high sucrose content was sub-optimal for single particle cryo-EM, therefore a buffer exchange and concentration step using a centrifugal filter was necessary before the sample was applied to glow-discharged EM grids.

For the *Chaetomium thermophilem* ct90S cryo-EM sample the ectopic integration of affinity tagged proteins required the development of a selection method, since conventional selection markers are unstable at the optimal growth conditions of the thermophile fungus. A thermostable resistance marker terbinafine was found to be appropriate for selection of positively transformed colonies (Kellner et al., 2016). The C-terminally FTpA-tagged ctUtp6 purification yielded a highly pure ct90S sample, without a free UTP-B sub-complex Utp6 is part of. It was therefore possible to directly use the eluate for cryo-EM without an additional sucrose gradient step followed by a buffer exchange step as in the case for sc90S.

These samples were then analyzed using negative stain EM and cryo-EM in our laboratory. For cryo-EM the samples were vitrified and data were collected on the 120 kV Spirit microscope by Charlotte Ungewickel and later Susanne Rieder. The high-resolution data collection on the 300 kV Titan microscope was performed by Dr. Otto Berninghausen. Negative-stain images of UTP-A, UTP-B, 5'ETS and sc90S, were kindly provided for analysis by Dr. Dirk Flemming (Hurt Lab).

My contribution was data analysis, image processing, initial model reconstruction, projection-matching refinement, structural interpretation and model fitting for the entire sc90S and ct90S projects and related samples (UTP-A, 5'ETS). The model building effort was facilitated by Dr. Jingdong Cheng (Beckmann Lab) who built the U3 snoRNA model and identified several proteins in the structure.

Negative-Stain Electron Microscopy

Negative-stain images of UTP-A, UTP-B, 5'ETS and sc90S were recorded by Dr. Dirk Flemming (Hurt Lab) who kindly provided boxed out particles for analysis. Freshly prepared affinity-purified samples were stained with 2% uranyl acetate following standard staining protocols. These images were collected on a Tecnai F20 transmission electron microscope (TEM), operated at 200 kV with an Eagle 4k CCD detector, calibrated to a pixel size of 3.821 Å/px at the specimen level. Particles were manually picked using the software BOXER (Ludtke et al., 1999). The total picked particles was 8629 for UTP-A, 5585 for UTP-B, 10198 for 5'ETS, 5466 for sc90S, and 6603 for ct90S.

Spirit cryo-EM

For cryo-EM of the *S. cerevisiae* 90S complex, 3 µL of freshly prepared sample was applied to Quantifoil R3/3 grids coated with 2 nm thick carbon layer and blotted for 3 s at 4 °C in a FEI Vitrobot MKIV. The data-collection was performed on a FEI Spirit microscope, operated at 120 kV and calibrated to a pixel size of 3.17 Å/px at the specimen level. The micrographs were recorded under low-dose conditions ($20 \text{ e}^-/\text{Å}^2$) in a defocus range from 1.0 to 4.0 µm. For the contrast transfer function (CTF) estimation CTFFIND3 (Mindell and Grigorieff, 2003) was used. The 400 collected micrographs were manually screened for drift and astigmatism and 150 micrographs with good power spectra were selected. 6000 particles were initially picked using XMIPP (Sorzano et al., 2004) and after manual inspection 2609 were kept for further processing.

Negative-Stain (NS) IMAGIC Workflow

3762 images of negative stained sc90S with a box size of 220 x 220 pixels were band-pass filtered between 8 and 500 Å and normalized to a zero mean value with a standard deviation (σ) of 2. Images were then iteratively centered to a rotationally averaged total sum of the data-set. Subsequently the translations and a circular soft mask with a radius of 50 pixels were applied. In the following multivariate statistical analysis (MSA) step the masked images were decomposed into 40 eigenimages or principle components in 64 iterations with the overcorrection factor set to 0.8. After the dimensionality reduction step the images were classified into several classes (2, 3, 4, 5, 6, 9, 16, 32, 64, 96, 110,

128) and visually inspected. A few 2D class averages were selected for the subsequent multi reference alignment (MRA) followed by another round of MSA.

ISAC Workflow

2D reference-free alignment of the negative stain data was performed in SPARX (Hohn et al., 2007). Before running **I**terative **S**table **A**lignment and **C**lustering (ISAC) images were usually first centered and down-sampled to 64 x 64 pixels (Yang et al., 2012b). In the case of the cryo-EM Spirit data, the RELION particles could not be exported with applying the contrast transfer function (CTF) correction. Therefore the defocus values were written into the header of the hdf particle stack using the SPARX inbuilt libraries in Python.

Cryo-EM of the *S. cerevisiae* 90S complex

For cryo-EM of the *S. cerevisiae* 90S complex, 3 μL of freshly prepared sample was applied to Quantifoil R3/3 grids coated with 2 nm thick carbon layer and blotted for 3 s at 4°C in a FEI Vitrobot MKIV. The data collection was performed on a FEI Titan Krios transmission electron microscope (TEM), operated at 300 kV, equipped with a TVIPS TemCam F816 CMOS detector and calibrated to a pixel size of 1.34 Å / px at the specimen level. The micrographs were recorded under low-dose conditions ($20\text{ e}^- / \text{Å}^2$) in a defocus range from 1.0 to 4.0 μm . For the contrast transfer function (CTF) estimation CTFFIND3 (Mindell and Grigorieff, 2003) was used. The 8809 collected micrographs were manually screened for drift and a maximum allowed 5% astigmatism cutoff was applied resulting in 5567 micrographs with optimal power spectra. After inspection for good areas 1492 micrographs were selected.

Particles were picked using the negative stain sc90S map as template for the **F**ast **L**ocal **C**orrelation (FLC) method implemented in **S**ystem for **P**rocessing of **I**mage **D**ata in **E**lectron **M**icroscopy and **R**elated fields (SPIDER) modular software package (Frank et al., 1981; Shaikh et al., 2008a). The 28,000 picked particles were analyzed using the classification-based verification method (Shaikh et al., 2008b) in SPIDER. Particles were grouped into 84 sets based on cross-correlation to the negative stain initial model. Subsequently within each set particles were classified in a reference-free manner. The

class averages for a particular projection group were then visually compared to the corresponding projection which made sorting out of ice artifacts and other contaminants fast and straightforward. The 21k particle coordinates that remained after screening were imported into **Regularized Likelihood Optimization** (RELION) software package (Scheres, 2012). The data-set was analyzed by 2D classification and 10 k particles from class averages with a low rotational average were selected for further processing. Using the initial model obtained from negative stain sc90S this data-set refined to the same 29 Å when filtering the starting model to 60, 80 or 100 Å. Additional particles were added by re-picking the micrographs with e2boxer (Tang et al., 2007). Similarly as before the particle coordinates were imported in RELION. The particles which overlapped with previously picked particles were discarded to avoid duplicate particles in the data. In the same manner as for the SPIDER picked particles, the data were 2D classified and refined. Plotting the maximum log likelihood ratio as a function of defocus a bimodal distribution was observed and the 6.7k upper mode particles were selected. The lower mode particles were also refined resulting in a considerably worse map than before the data-set split. The resulting gold-standard refinement of the upper mode data-set improved in resolution to 19 Å and with post-processing to around 14–15 Å depending on the masking parameters. According to RESMAP (Kucukelbir et al., 2014) the local resolution in the sc90S core was around 15 Å. A scheme of the sc90S cryo-EM processing can be found in Figure 10.

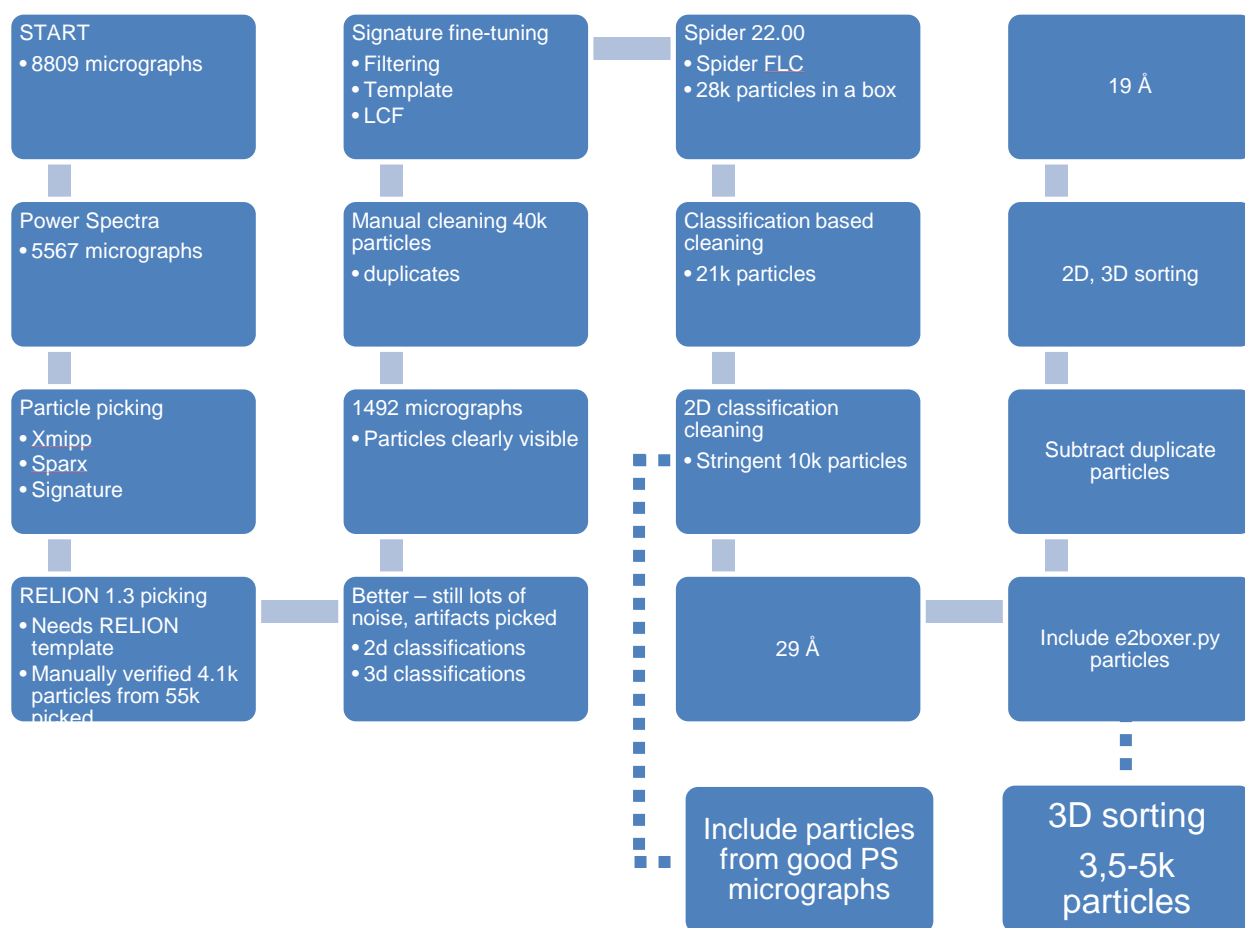


Figure 10. The processing scheme for the sc90S cryo-EM map.

Cryo-EM of the *C. thermophilum* 90S Complex

For cryo-EM of the *C. thermophilum* 90S complex, 3 μL of freshly prepared sample was applied to Quantifoil R3/3 grids coated with 2 nm thick carbon layer and blotted for 3 s at 4°C in a FEI Vitrobot MKIV. The data-collection was performed on a FEI Titan Krios TEM, operated at 300 kV with a Falcon II detector, calibrated to a pixel size of 1.062 Å/px at the specimen level. The micrographs were recorded in movie mode at a dose rate of 50 $\text{e}^-/\text{Å}^2/\text{s}$ in a defocus range from 1.0 to 3.5 μm . Movie frames were acquired and grouped into seven segments. First five frames were stored in the first five segments, the next three frames were summed in the sixth segment, and the next four frames in the seventh segment. The remaining collected frames were not stored. The final c90S map was reconstructed with the first 5 segments corresponding to a total

dose of 16 e⁻/Å². Whole-image drift correction of the frame segments was done with MotionCorr (Li et al., 2013) and CTFFIND3 was used for estimating CTF parameters. The 4800 collected micrographs were manually screened for drift and a maximum allowed 5 % astigmatism cutoff was applied resulting in 3725 micrographs with good power spectra. After inspecting for good areas 1428 micrographs remained.

Particles were picked using SIGNATURE (Chen and Grigorieff, 2007). The sc90S cryo-EM map was low-pass filtered and projections of characteristic different views were used for the picking template. To separate non-alignable particles and artifacts a multi-reference projection matching step was done in SPIDER using the sc90S cryo-EM as reference and a volume reconstructed from noisy particles with most density concentrated at the edge of the box. From the initial 88k particles 81k particles remained after this iterative projection matching step and the particle coordinates were imported into RELION. These images were then 2D classified and the particles from good classes were kept. The final map was refined from 43k particles to an average resolution of 7.3 Å using the gold-standard refinement.

Cryo-EM Model Creation

I have fitted the coordinates of the 18 out of the 19 found WD-40 repeat containing proteins in the ct90S structure and used the crystal structure of the *S. cerevisiae* 80S ribosome (PDB code 4V7R) (Ben-Shem et al., 2010) as a starting point for modelling the pre-18S rRNA. Regions of the 18S which were identified in the ct90S map were rigid-body fitted into the density in Chimera (Pettersen et al., 2004) and adjusted in Coot (Emsley et al., 2010). The ribosomal proteins (r-proteins): uS4, uS7, uS8, uS9, uS11, uS15, uS28 were rigid-body fitted from the *S. cerevisiae* 80S (PDB code 4V7R) and r-proteins: eS1, eS4, eS6, eS7, eS8, uS17, eS24 were rigid-body fitted from *T. thermophila* 40S (PDB code 4V5O) (Rabl et al., 2011) crystal structures. The WD-40 repeat proteins were rigid-body fitted in Chimera using part of the crystal structure containing one β-propeller fold of Utp21 (PDB code 4NSX; residues 20 to 32 and 348 to 659) (Zhang et al., 2014) as template. The exception was Rrp9 (PDB code 4J0X), whose identity was determined from the known association to the U3 snoRNA (Zhang et al., 2013). For fitting Imp3, the *S. cerevisiae* homolog uS4 (PDB code 4V7R) was used.

Imp4 was fitted using the Brix domain of Rpf2 taken from the crystal structure of the Rpf2-Rrs1 complex (PDB code 5A53) (Madru et al., 2015). For the Rcl1-Bms1 complex, the crystal structure of Rcl1 in complex with a short Bms1 peptide part (PDB code 4CLQ) was used (Delprato et al., 2014). The U3 snoRNP proteins (Nop1, Snu13, Nop56, Nop58) were fitted using the crystal structure (PDB code 3PLA) from *S. solfataricus* (Lin et al., 2011). Nop1 and the N-terminus of Nop56 and Nop58 were adjusted manually to fit the density using Coot. The *ct*U3 snoRNA (nucleotides 64-105, 108-117, 126-206 and 239-270) was modelled *de novo* in Coot. For the KH-1 and KH-2 domains, the crystal structure (PDB code 4QMF) was used (Zheng et al., 2014). For fitting the Kre33, the crystal structure of the *E. coli* homolog was used (PDB code 2ZPA) (Chimnaronk et al., 2009). For fitting the Emg1 from the *S. cerevisiae* the crystal structure (PDB code 3OIJ) was used (Thomas et al., 2011). For fitting Utp30, the *T. thermophilus* homolog crystal structure (PDB code 2HW8) was used (Tishchenko et al., 2006). For fitting Utp24, the *S. cerevisiae* homolog crystal structure (PDB code 4MJ7) was used (Lu et al., 2013). Part of Utp10 predicted to contain α -helices which adopt an α -solenoid fold was modeled using the PHYRE2 web-server (Kelley et al., 2015) with the structure of Importin beta-3 (PDB code 3W3Z) as template and manually adjusted in Chimera to fit the density. The three long α -helices, Utp6 and the oligomerization domains of UTP-A and UTP-B sub-complexes were made in Chimera using poly-Ala residues and applying α -helical constraints.

Figure Preparation

All the images showing molecular models and 3D electron densities I have generated with UCSF Chimera (Pettersen et al., 2004).

Results

Sc90S Processing

To obtain an initial view of the yeast 90S complex the first step was to cluster similar images together into 2D class averages with an improved signal-to-noise ratio (Frank, 2006a). The negatively stained images were clustered using different methods following the simple rationale that if the 2D class averages obtained from analyzing the same data-set using different software packages would show similarities in their features, the data contain enough signal to allow reliable alignment and subsequent clustering of these images into groups of similar particles. Only after this prerequisite has been ascertained the data were considered amenable for generating an initial model and for further processing.

Since we were dealing with an unknown sample, to increase the internal coherence establishing a solid understanding of the data was considered paramount. One has to also consider that simple parameters such as the radius of the particle had to be first approximated and then optimized to obtain best results. In the end there was not one key step or software package that made all the difference, but the sum of optimizing several steps of the image processing workflow.

Another aspect worth considering was the unknown amount of heterogeneity in the data. A side-benefit of subjecting the same data-set to analysis from different software packages are the subtle differences that can be obtained as a result of employing different algorithms or their implementation. Therefore comparison of the results in a consensus approach provides an opportunity to discriminate between the main features representing the bulk of the data and features in subpopulations which have to be considered with caution.

As the project progressed in addition to obtaining characteristic features of the particles in different orientations, the sample preparation also improved and from approximately 5

particles per micrograph for the initial sc90S micrographs the number improved to around 20 particles for the ct90S that in addition appeared more homogenous.

Negative-Stain (NS) – 2D Classification

The sc90S negative stain data were used to reconstruct an initial model that was subsequently used for the cryo-EM sc90S projection-matching refinement. The negatively stained particles were first classified into 2D class averages using different software packages. Two reference-free methods were used implemented in Sparx and RELION (Scheres, 2012) as well as a user guided classification method implemented in Imagic.

NS – ISAC Results

After performing Iterative Stable Alignment and Clustering (ISAC) 65 % of the whole data-set ended in 21 stable class averages after the first generation of the ISAC run revealing a globular particle with overall dimensions of 36 by 41 nm (Figure 11). These class averages represent the bulk of the data, but also show that a considerable percentage of the data was unclassifiable, indicating heterogeneity within the data.

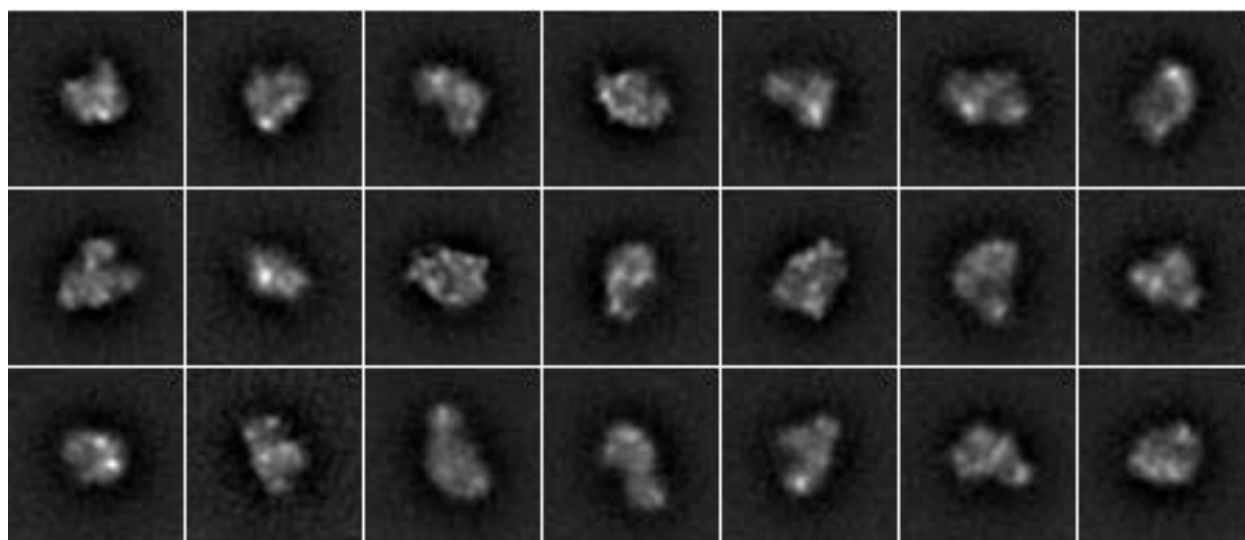


Figure 11. ISAC 2D class averages showing different sizes particles present in the population. A typical 90S particle very much resembling the later obtained cryo-EM densities can be seen in the first column of the second row. The white dots in the particle are most likely staining artifacts that produce the appearance of density rich regions through averaging.

NS – RELION Results

The same data-set was then classified again using a regularized maximum likelihood (ML) based reference-free classification method implemented in RELION where the data was classified into $K=50$ classes (Figure 12).

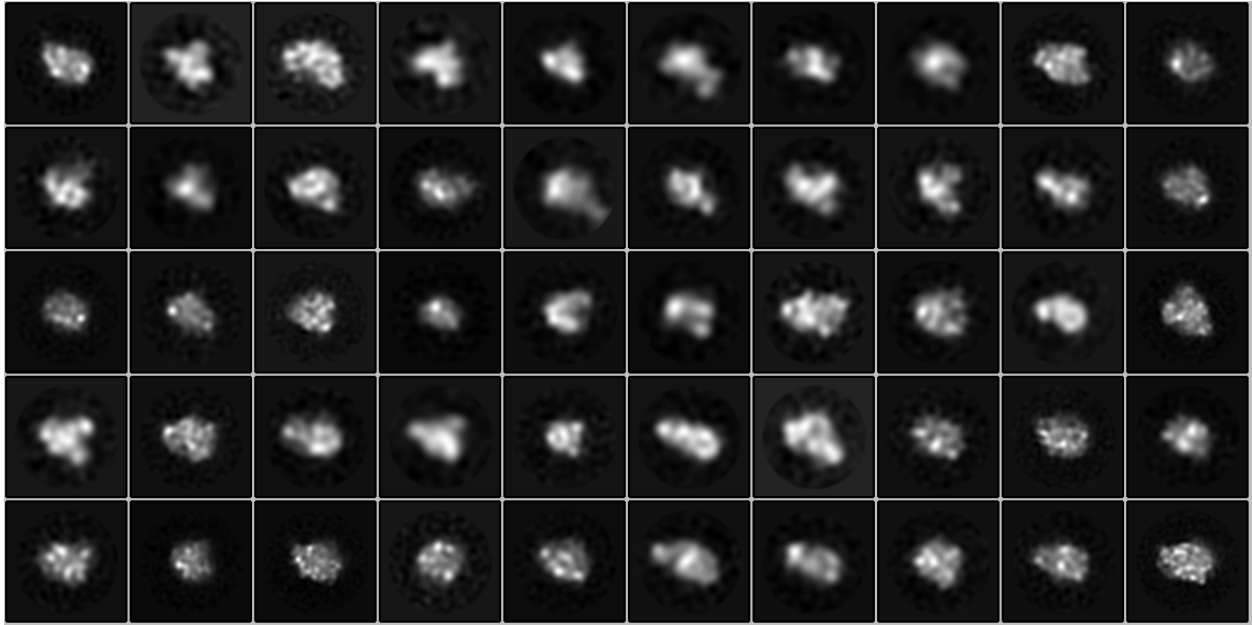


Figure 12. RELION 2D class averages. The same typical 90S particle as obtained with the ISAC classification can be seen in the last column of the third row. Many fuzzy class averages can be seen as a result of the whole data being classified.

Class averages with the lowest rotational average, highest homogeneity or i.e. the best classes were similar to the results obtained from the ISAC classification run. The differences in the classes with fewer particles are resulting from different methodologies of the methods, mainly since in RELION all the particles are classified once using K-means as opposed to ISAC where the K-means classification is iteratively repeated and the results of different classification runs are compared to yield repeatedly occurring stable class averages and the unstable particles are not forced into classes. The regularized ML mitigates to a point the shortcomings of K-means clustering.

NS – IMAGIC Results

Third, the same data-set was analyzed with multi-variate statistical analysis (MSA) on centered particles followed by a user biased multi reference alignment (MRA) step in IMAGIC (van Heel et al., 1996) (Figure 13). Picking particles from the micrographs offers the benefit of being able to visually verify if the obtained 2D class averages resemble some of the picked particles, on the other hand the user is biased in his decisions.

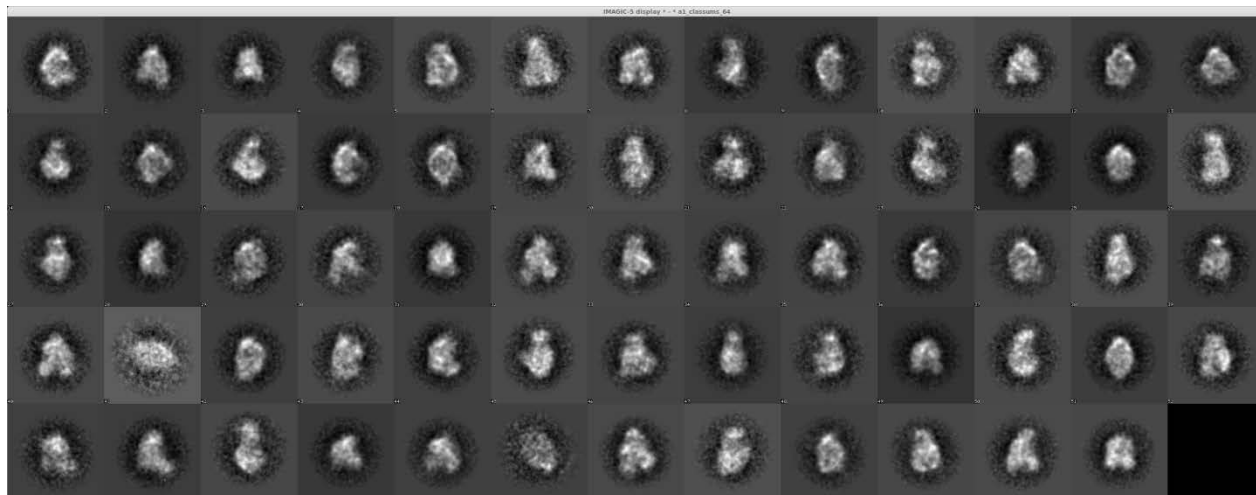


Figure 13. User guided 2D classification in Imagic shows similar class averages to other methods. The class averages are less sharp, but the typical 90S view can be seen in several class averages, for example, in the first column fourth row.

NS – Comparison of the 2D Class Averages

The comparison of the obtained 2D class averages from all three software packages showed classes with similar features indicating the data were classifiable and in turn averageable. In addition, during micrograph inspection particles were observed which resembled the obtained 2D class averages. Subtle differences were observed in a few class averages (Figure 14), which prompted the question of the degree of heterogeneity in the sample.

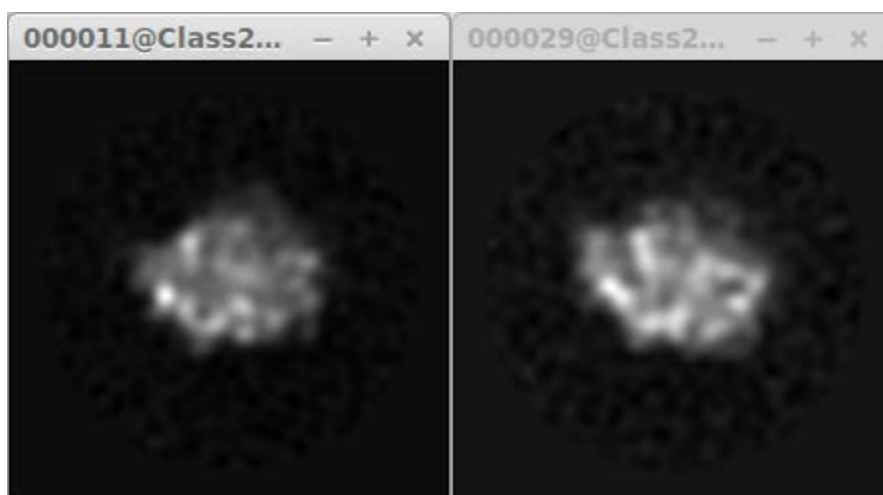


Figure 14. RELION 2D class averages exhibiting small differences.

NS – Maskiton

Maskiton was then used to further classify the particles within individual class averages (Yoshioka et al., 2013). However, since we had no prior structural knowledge of the particle a meaningful interpretation of the obtained self-organizing maps (SOM) was not possible rendering the discovery driven approach meaningless.

NS – Initial Model Creation

Next step was generation of a number of initial models using predominantly Simple Prime (Elmlund et al., 2013) from the obtained 2D class averages from Isac Sparx and RELION and derived sub-sets. Eman2 was also used for reconstructing some initial models which turned out very different between runs on the same data-set. Therefore, after looking at the Python script how the initial models were generated Eman2 was no longer used, mainly because there was no selection criteria available, which would discriminate between initial models.

What made sense at the time was using the ISAC 2D class averages as a starting point for Simple Prime. The unknown heterogeneity of the sample was problematic and the rationale was to constrain the stochastic hill climbing approach to the pre-selected ISAC classes as well as particles composing them. The 2D class averages from MRA in Imagic were also used as input for the *ab initio* model generation. This resulted in a set

of Simple Prime models, which were used as initial models for subsequent refinement in RELION.

Spirit cryo-EM

In parallel to the negative stain image analysis we collected cryo-EM images on our screening microscope at 120 kV with the aim of obtaining another reference for comparison and to be able to estimate the degree of potential artifacts introduced by the heavy salt staining. Deformations of the structure by negative staining are not uncommon (Frank, 2006b). Another set of aims was to test the viability of the sample for flash-freezing, to optimize the parameters for particles picking, and to generate a cryo-EM initial model before moving to the next step of collecting data on a high-end microscope with higher acceleration voltage and consequently lower contrast in the micrographs.

From the collected 400 micrographs 150 had acceptable power spectra. The signal-to-noise ratio in the micrographs was sufficient for particle picking to work with XMIPP, Starfish and Signature. However, in the 2D and 3D classifications that followed, reproducibility of obtaining similar results was poor, and did not improve after manual particle inspection resulted in a cleaner data-set containing 2609 particles.

RELION 2D classes could be obtained which resembled to a certain degree the 2D class averages from negative stain. ISAC did not find any stable class averages unless the threshold level was lowered to an extremely low value (Figure 15). These class averages then resembled somewhat the negative stain counterparts, but revealed a perforated particle, which didn't resemble at all the expectations of a biological macromolecular complex.

If the resolution could be increased to recognizing some structural features, this would be a confirmation that we are being on track. The Imagic MRA class averages showed no particular features that would help visually relate the class averages to each other.

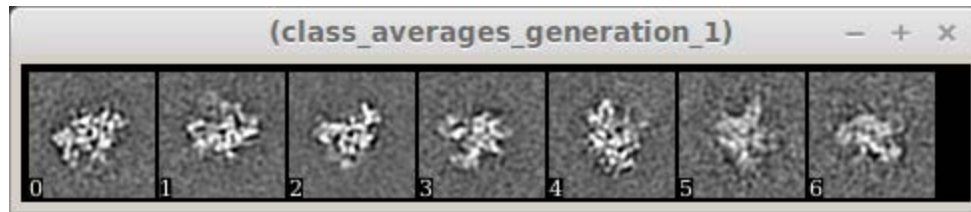


Figure 15. 2D class averages obtained after considerable loosening of the threshold criteria in ISAC. The images ending up in the 6 class averages represent only a small fraction of the data-set, but they show similar features to the negative stain classification results.

Spirit cryo-EM – Initial Model Generation

The 2D classes had a perforated appearance and it was difficult to imagine how they would add up in 3D. After some initial trial and error I've decided for a brute force approach using Imagic. The rationale was to guess a close enough starting model to which the particles even with a poor signal-to-noise ratio would align in projection matching.

To that end I wrote a Python script for permutating different sub-sets of class averages whose relative orientations were then estimated using common lines in Imagic. Several models were generated for a given n - and k - and the models were scored based on the correlation of the 2D class averages to the obtained initial model. The best models of a particular run were visually inspected, but models with similar features were not observed.

NS – RELION 3D Classification

Since 3D classification in RELION of the negative stain data of my first project on the RSC remodeler showed two distinct sub-populations of the particles I used the same approach for the 90S data. Varying the number of classes from 3 to 8 resulted in one major sub-population repeatedly comprising a third of the data-set. In addition, an overall similar model was reconstructed in an independent reconstruction attempt in RELION using a subset of particles with low rotational averages after the 2D classification and refinement procedures. The consistency in the 3D classifications and the independent reconstruction were a good indication that this is not just a coincidence.

Visual comparison of different projection views (Figure 16) to the boxed out particles was done and similarities were observed. From certain angles the map had holes at a reasonable contour level and such features were observed in the boxed out particles.

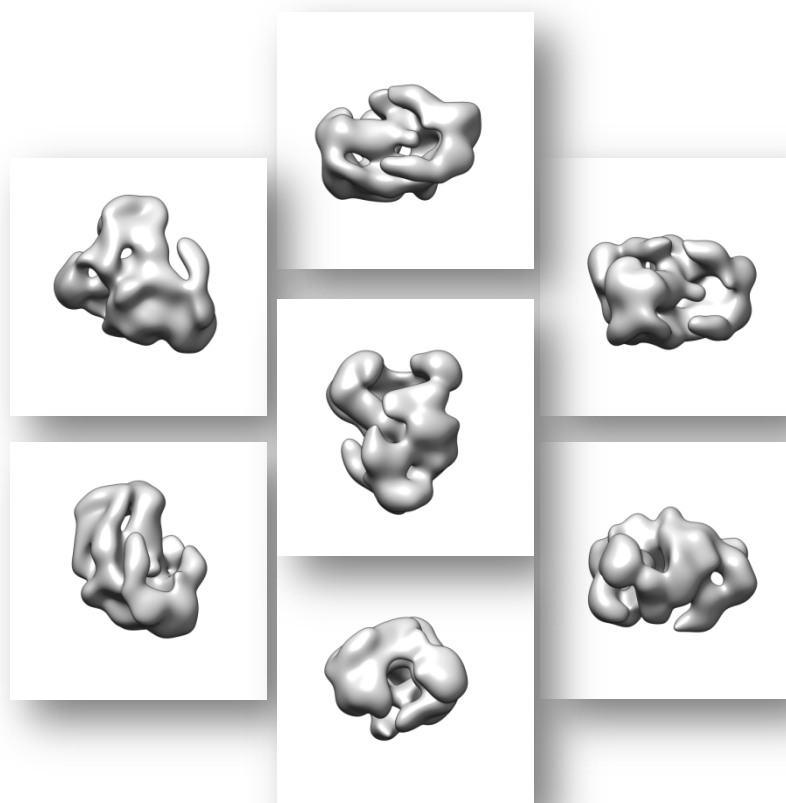


Figure 16. Different views of the obtained initial model from negative stained sc90S images are shown next to the central front view of the map. The hollow appearance of the projections and the apparent holes penetrating the particle from certain angles were unexpected.

NS – Validation of the Models Spider

The main reason why the model was treated with caution was the hollow appearance which would be very unusual for a biological macromolecule. There was a typical 2D class average view common to the negative stain and Spirit cryo-EM images that matched the shape of the candidate initial model. In order to investigate this further in a more quantitative manner I've compared how the projections of the candidate models

correlate with the obtained 2D class averages using Spider (the details how this was done can be found in the methods section) (Figure 17).

Figure 17. Initial model validation. The 2D class averages are shown in the even numbered images. The projections with the highest cross-correlation are shown right of the 2D class averages in the odd numbered images. If a mirrored projection had better correlation the letter M stands after the projection number. The gallery displayed is not sorted. The matching pairs of 2D class averages and projections are highlighted in green.

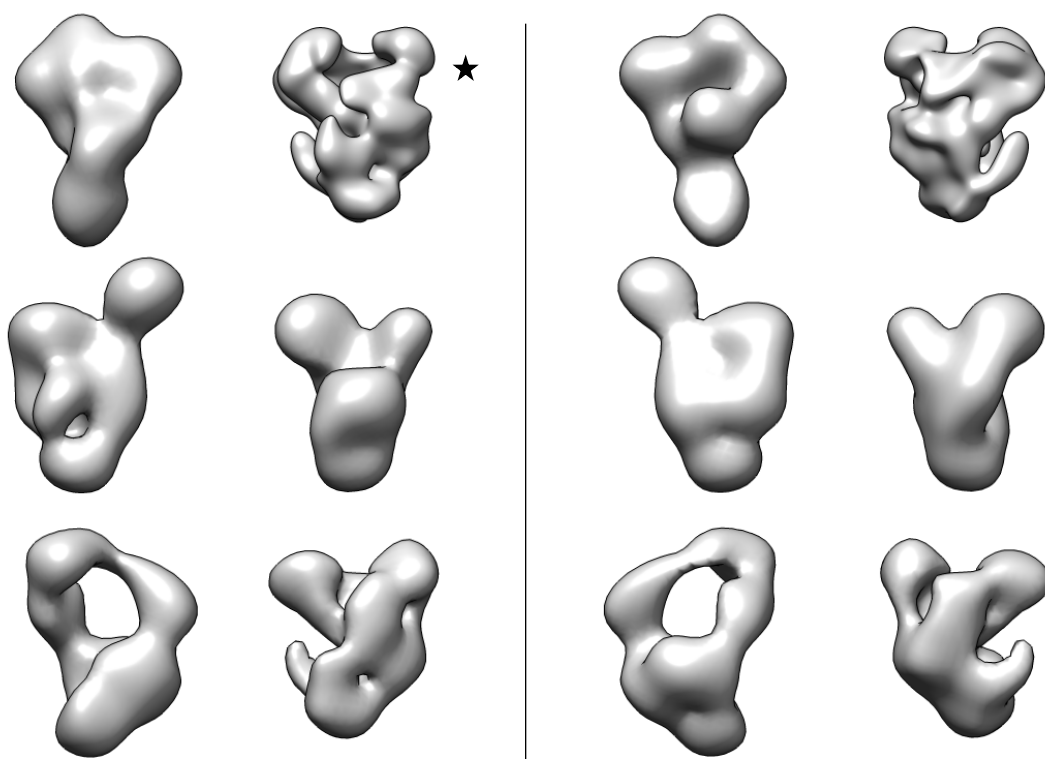


Figure 18. RELION 3D classification results from the whole data-set of 3,7k particles into 6 classes. The initial model indicated by the star represents 32 % of the data-set and had the best overall score to the reference-free 2D class averages. Front views (left) and back views (right) separated by a vertical line.

NS – Tilt-Pair Analysis

To determine the handedness of the initial map tilt pairs were collected. The approximately 20 particles from 5 micrograph pairs showed a scattered plot without a clear indication of the hand (Wasilewski and Rosenthal, 2014). The correlation to the projections of the initial model was also very poor except in one instance (Figure 19), where the hand would have been the mirrored of the initial map. This turned out to be the case when the higher resolution of the sc90S cryo-EM maps revealed the dsRNA helices.

Particle 2
Scatter plot

Minimum phase residual 69.9°
Minimum at position (-12.0°, -5.0°)
Tilt angle 13.0°
Hand difference 10.8°

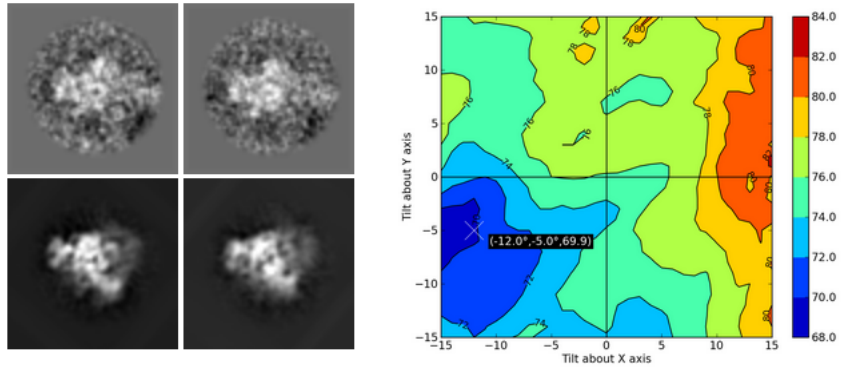


Figure 19. Tilt pair analysis of the negative stain data showing the tilted particles (gray boxes) and corresponding projections of the initial model (dark gray boxes). The blue region in the heat map on the right indicates the mirrored map would be a better fit.

Titan cryo-EM sc90S

With the 300 kV Titan cryo-EM data the particle picking which was not a problem for the 120 kV Spirit cryo-EM data became one. Particle picking methods that worked previously failed due to poor signal-to-noise ratio in the micrographs. Given the poor micrograph quality I've decided the best chance of success lies in using a template-based approach (Moody, 2011) (Figure 20).

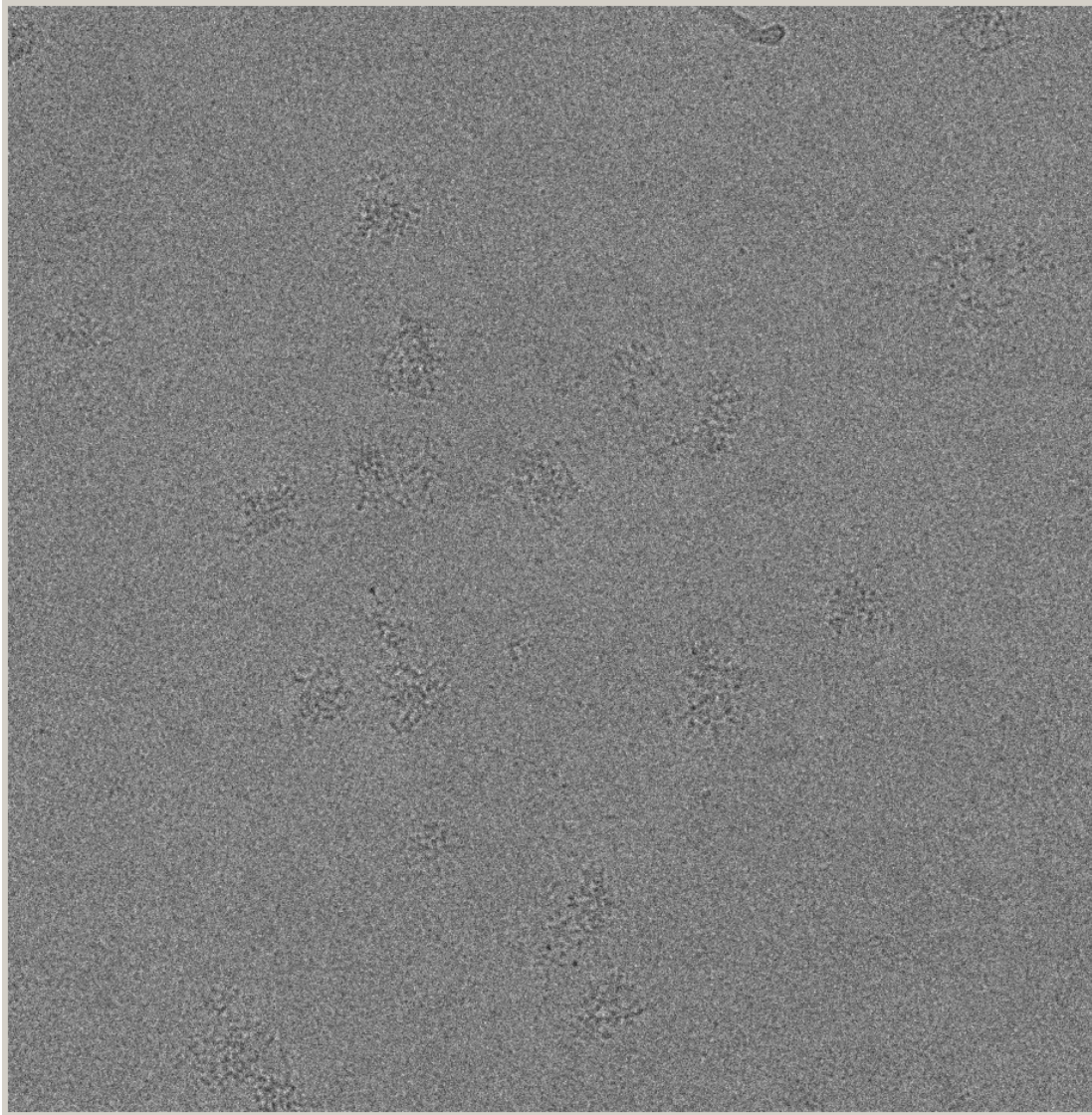


Figure 20. One of the better cryo-EM micrograph of the sc90S sample where particles were visible.

The template from the negative stain initial model did not produce great results with Signature, even after extensive parameter optimization. Particles could be found for a given set of LCF and SCF parameters in a narrow defocus range. High degree of sensitivity for small parameter changes as well as a high defocus dependence in finding particles in micrographs was unusual and indicated there might be a problem with the template. Adjusting the parameters for a particular defocus range improved the particle picking and partially mitigated the problem. Instead of linearly interpolating between the parameters as a function of defocus I decided to check whether or not the particle size given the estimated pixel sizes of the negative stain initial model and the cryo-EM images is the same. There could be artifacts introduced by staining or size differences due to not optimally determined pixel size, since the data for negative stain and cryo were collected on two different microscopes. Changing the pixel size of the initial model in small increments using a script showed the best cross-correlation of several hundred manually picked cryo-EM particles to the reference was by reducing the negative stain pixel size by 9%. After applying this correction the template-based particle picking improved considerably.

In parallel, template-based picking from 2D classes of manually picked cryo-EM images in RELION didn't find centered particles after extensive optimization of parameters and re-running picking on the calculated figure of merit (FOM) maps. Duplicate particles were also occasionally observed by visual inspection of the picked particle stack. Changing the spacing parameters between particles did not seem to resolve this problem adequately.

Finally, the best results were obtained using a template-based method implemented in a Spider protocol for Fast Local Correlation (FLC) (Shaikh et al., 2008b). This method found most particles in the images. Particle numbers varied from 5 to in very rare cases 20 particles per micrograph and it was essential to pick most of the few particles to have at least a minimal amount of particles to work with and reach a resolution where secondary structure features would become recognizable. Setting the threshold relatively low selected most particles, but at the same time a lot of non-particles, ice contaminations and other unwanted artifacts were picked. The advantage was that in

subsequent steps the falsely picked particles could be selected out in a classification-based verification approach (Shaikh et al., 2008b). From initially 28k picked particles 21k particles were selected after screening.

Titan cryo-EM – 2D Classification

The obtained centered particle coordinates were imported from Spider into RELION. There the particles underwent 2D classification and the class averages with sharpest features were selected resulting in 10k particles (Figure 21). These class averages were still not great compared to results from typical 80S or 70S ribosomes, but considering the low particle numbers and sub-optimal data quality these particles were by far the best sc90S particles to date.

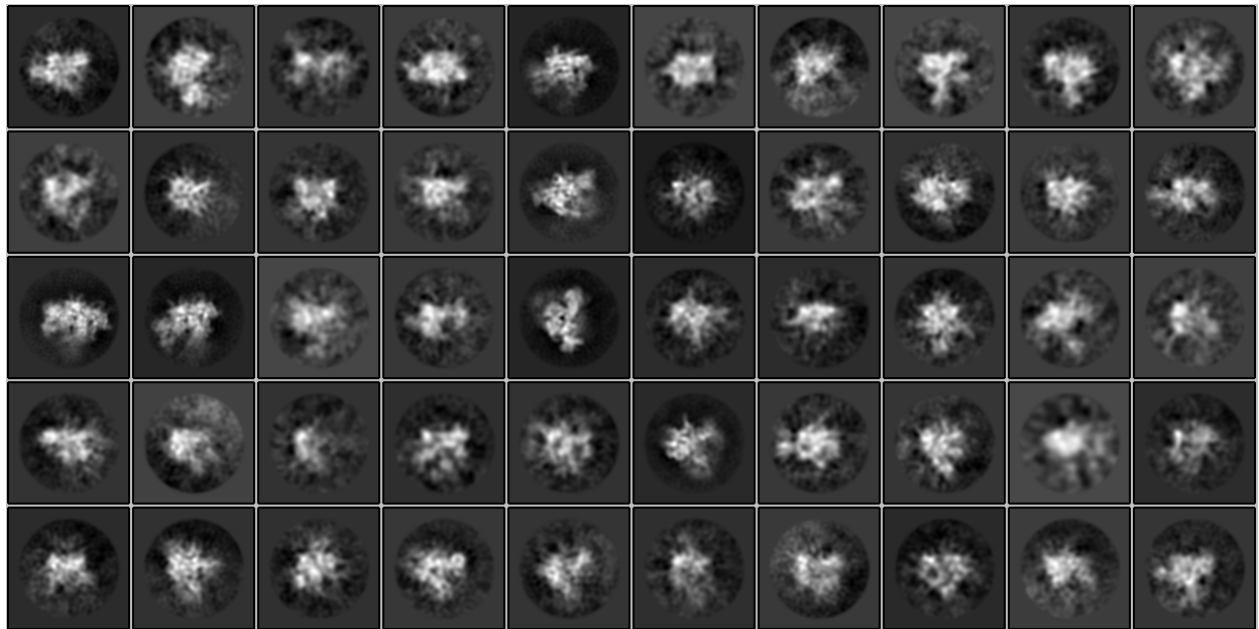


Figure 21. RELION 2D class averages of the 10k particle sub-set. The typical 90S particle can be seen in the fifth column third row, showing more detailed features than the negative stain data. However, on the top left of the particle, the density appears less defined compared to the previously obtained negative stain 2D class averages.

Titan cryo-EM – 29 Å Refined Map

This sub-set of 10k particles refined to 29 Å using the obtained negative stain initial model. The refined map contained density in regions the initial model had none,

therefore a complete reference bias was excluded. The overall same 29 Å map was obtained after filtering the initial model to 60, 80 and 100 Å indicating that this data-subset converges to a similar result even when relaxation of the starting map is relatively high. The resolution of the obtained map was not in the region where WD-40 repeat containing proteins (β -propeller fold) or double stranded RNA (dsRNA) helices could be recognized. It became apparent that the particle was indeed perforated in appearance and not as compact as mature ribosomes (Figure 22).

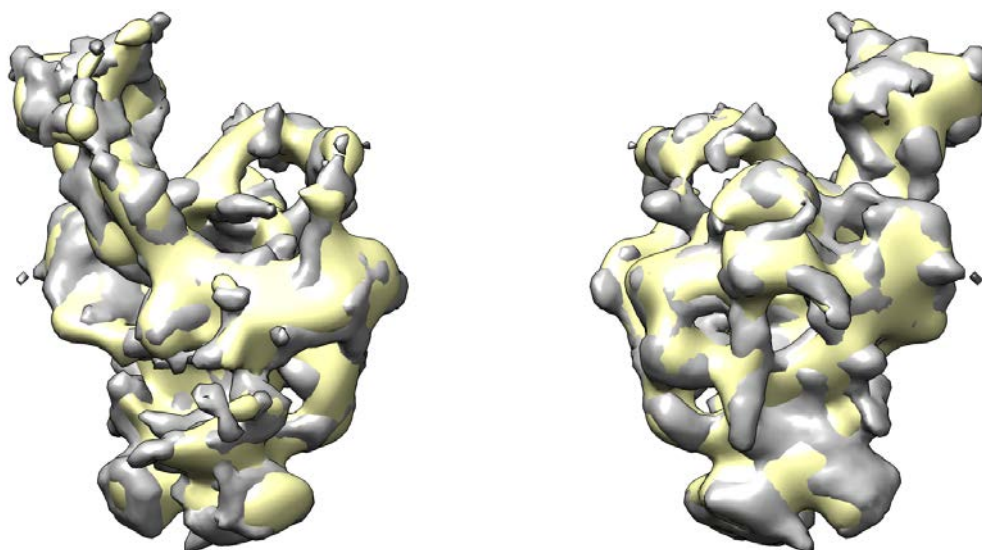


Figure 22. Overlay of the negative stain initial model (yellow) and the refined cryo-EM map showing additional density outside the initial model and an overall exposed and accessible core as opposed to the compacted structure of a mature ribosome.

Titan cryoEM – 19 Å Refined Map

To improve the resolution additional particles were added to the Spider picked particles with e2boxer. Their coordinates were imported into RELION and duplicate particles were removed. The processing was re-done and a bimodal distribution observed when plotting the maximum log likelihood value as a function of defocus. Refining the 6,7k upper mode particles improved the resolution to 19 Å with splitting the data in two halves during refinement. According to Resmap the local resolution at the core was around 15 Å (Kucukelbir et al., 2014) (Figure 23).

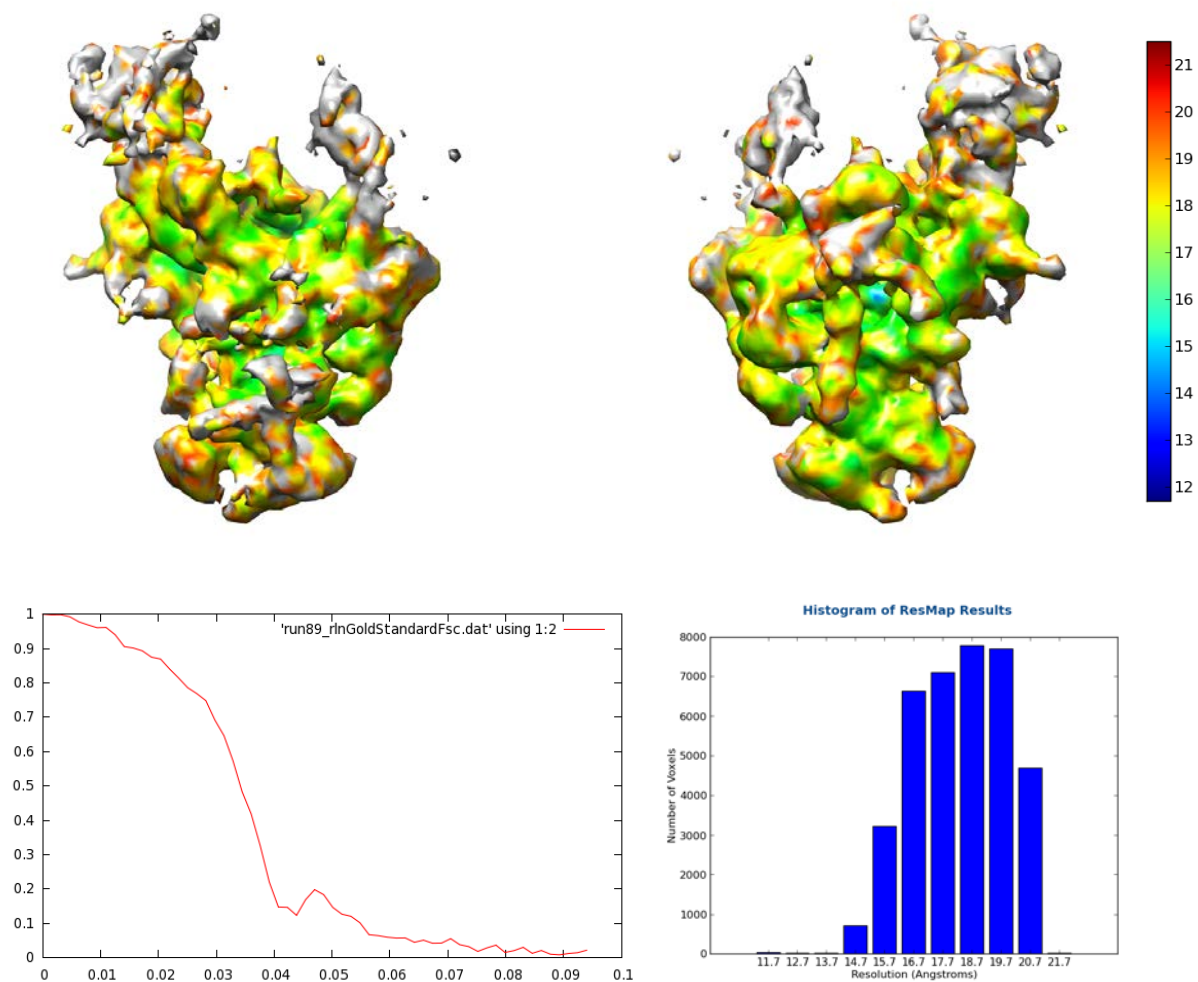


Figure 23: Top row: Local resolution map of the sc90S. Bottom row left to right: the FSC curve and a histogram showing the local resolution i.e. voxel distribution in bins.

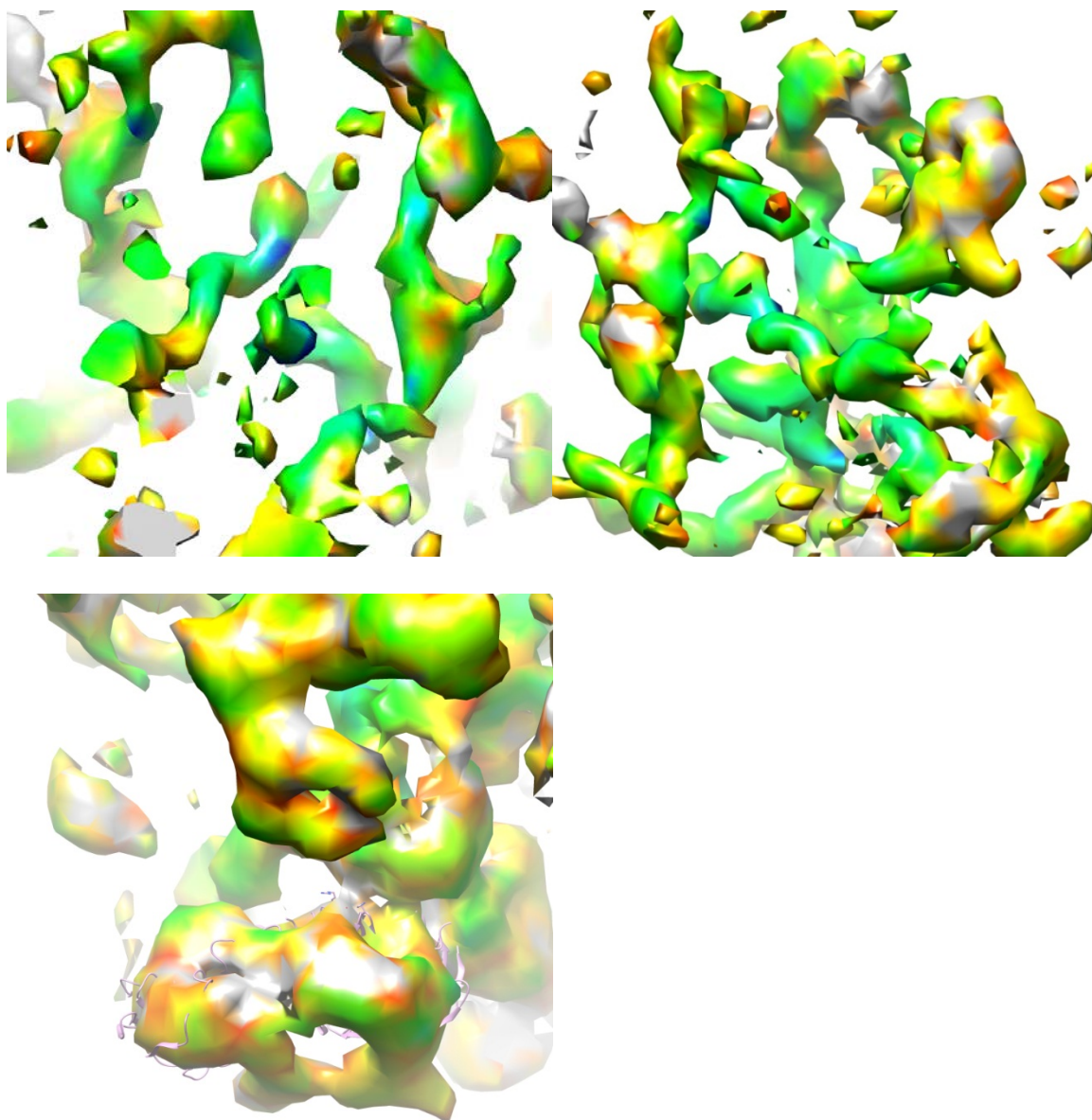


Figure 24. Top row: the core of the sc90S viewed at a high contour level reveals long spiral tubes resembling the pitch of the RNA form A helices. Bottom row: several ring-shaped densities were visible in the lower part of the map indicating a cluster of β -propellers. The crystal structure of the Utp21 is shown fitted into two adjacent ring densities.

Features in the core revealed spiral shaped rod-like densities typical of dsRNA and larger ring-shaped densities located mainly at the periphery typical of β -propeller proteins (Figure 24). The protein predictions for UTP-A and UTP-B proteins suggested

these two sub-complexes contain many beta-propellers. Fitting of the crystal structure of the Utp21 protein showed these ring-shaped densities can indeed accommodate a double β -propeller. Appearance of these structural features implied the obtained map was not a mere artifact of processing. In addition, the pitch of the rod-like density was the inverse of that of a typical dsRNA helix. This suggested that the refined map was in fact in the wrong mirrored world.

The search for structural features characteristic of the 18S rRNA, an integral part of the 90S pre-ribosome, started. The resolution of the map was not high enough to uniquely position the 18S rRNA. The density was oval from one side and looked at a right angle rotation narrow and could have accommodated an 18S in the upright orientation. Parts of the 18S fitted into the density, some parts did not. At this stage it was difficult to estimate to what extent we can find the 18S already resembling the mature state. Further exploring positioning of the 18S in different orientations in the sc90S map showed there was a fit which indicated that the 18S body would fit in the upside-down orientation equally well if not better as in the upright orientation. This turned out some 8 months later by obtaining one of the first 9 Å c90S maps to actually be the case.

Titan cryo-EM – RELION 3D Classification

3D classification of the 10k particles resulted in one class with stronger density in the upper part of the particle, but since the particle numbers were fewer the overall resolution was worse as compared to the refined map from the 6.7k particles, for example the pitch of the dsRNA helices at the core was no longer discernible. Surprisingly, a WD-40 repeat domain protein disappeared from this class even at low contour levels, suggesting it is a single and not a double β -propeller protein.

Ct90S pre-Ribosome

In order to further improve the resolution, the 90S pre-ribosome was isolated from a thermophilic fungus *Chaetomium thermophilum* (ct90S). The rationale was that the thermophilic complexes would at the low vitrification temperature be more rigid than, for example, 90S pre-ribosomes from yeast. This would be an advantage in averaging images, which is at the heart of the single particle cryo-EM. In addition, for ct90S data collection a direct electron detector was available omitting the scintillator step inherent signal loss, resulting in vastly improved signal-to-noise ratio in the micrographs in spite of sub-optimal ice thickness (Figure 25). In addition, the sample preparation was also improved and a fortunate species related difference resulted in isolated ct90S complexes without free sub-complexes as was the case for the yeast pullouts, which enabled the direct use of the eluate for making grids. Omitting the sucrose gradient step increasing the yield, which resulted in an 3 to 4 fold increase in particle numbers, and in addition caused less particles to dissociate and thereby the homogeneity of the particles seen on the micrographs was considerably improved.

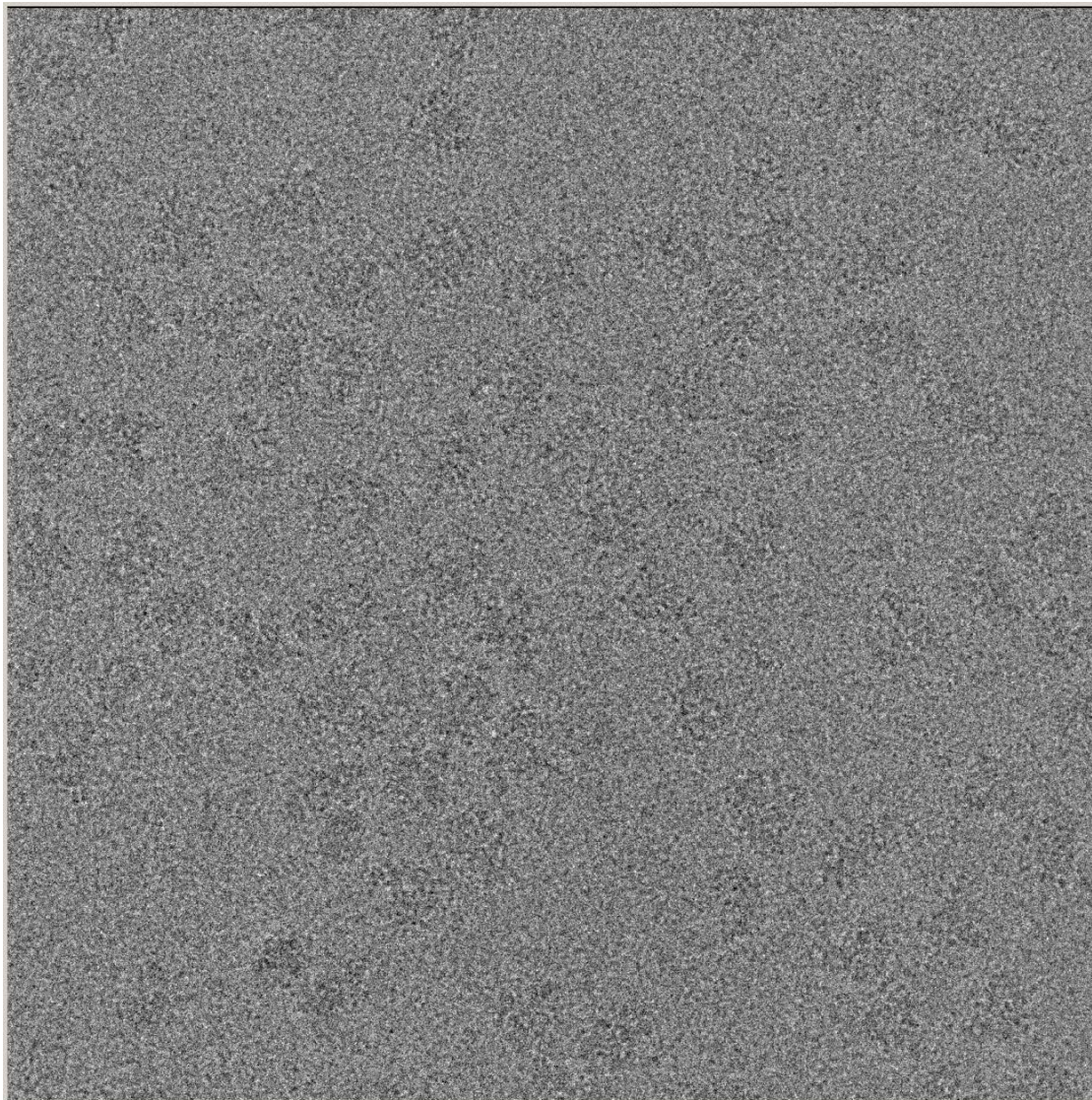


Figure 25. One of the better *ct90S* micrographs showing considerably more particles than in the *sc90S* micrographs.

Ct90S 3D Map at 7.3 Å Average Resolution

After the huge effort of obtaining a reliable *sc90S* cryo-EM the *ct90S* image processing was merely a projection-matching routine, that quickly resulted in a first 9 Å *ct90S* that was subsequently improved to 7.3 Å average resolution. The obtained *ct90S* cryo-EM density had overall dimensions of 360 by 400 Å at a contour level of 0.025. The particle appears to have an overall perforated structure, with a dense core surrounded by large empty cavities and considerable mass located at the periphery. Since it is a novel

structure, in order to describe it a common point of reference had to be established. We termed the top of the particle the head, the central part the body and the bottom part the foot or base (Figure 26). The head contains long tentacle like densities wrapping around the core and extending to the back side of the body. The body contains at its center a dense core that is extending from the head to the foot of the particle. From the core there are several connection towards the periphery, which builds a ring like scaffold all the way around the body, the region between appears to be rather density free giving space to large cavities and contributing to the impression of a scaffold supporting the central core. This scaffold of the body part contains at its periphery ten ring like densities which are found often in proximity to rod like densities, which have the same pitch than that of dsRNA helices. One such dsRNA is held in check by three such β -propellers, isolating it from the rest of the structure and rigidly pointing its tip at an angle away from the core. Finally the foot of the particle contains seven β -propellers and three dsRNA rods pointing in the same direction. An astonishing feature of the foot is a 250 Å long arm density connecting the foot to the head of the particle that can be observed at a lower contour level of 0.015 (Figure 27). Perhaps the most remarkable part of this long arm is that it is anchored at two points that constrain its rotational freedom and point it in an arch bow trajectory towards the head.

The local resolution of the particle assessed by Resmap is best at its core and radially deteriorates towards the periphery coinciding with the overall appearance of the density.

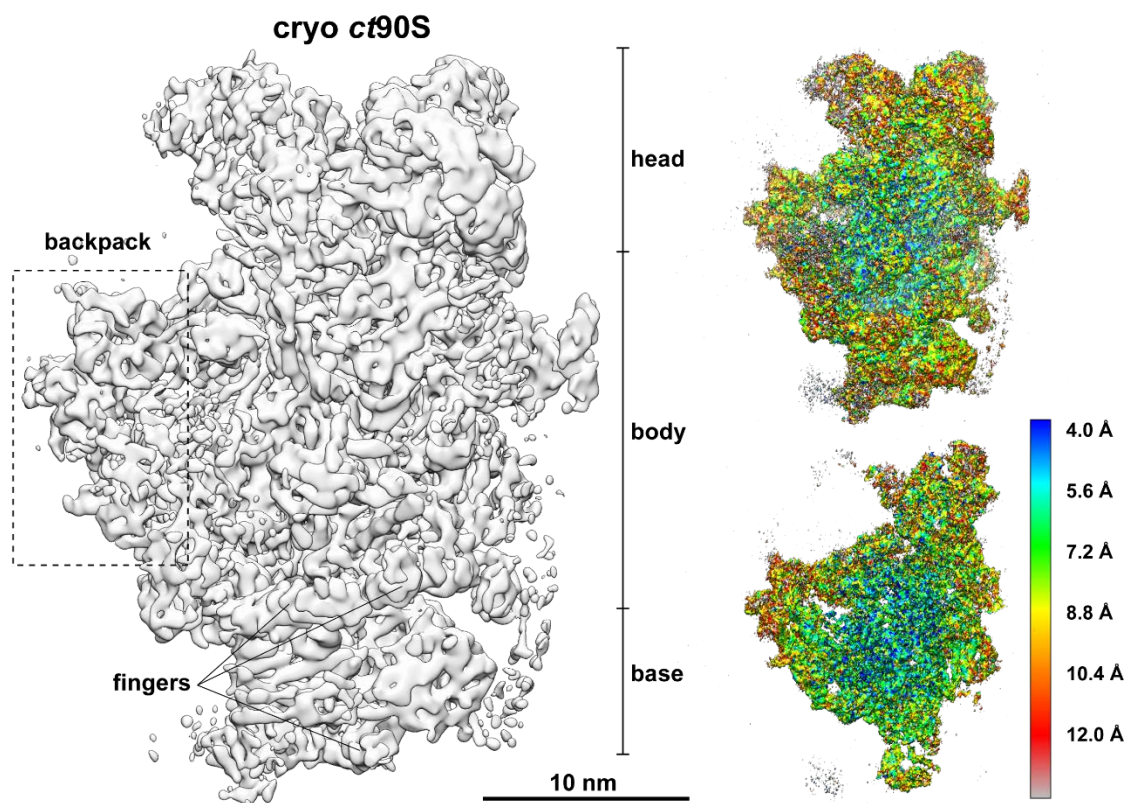


Figure 26. Left: Front view of the *sc*90S pre-ribosome. Right: Local resolution assessment.

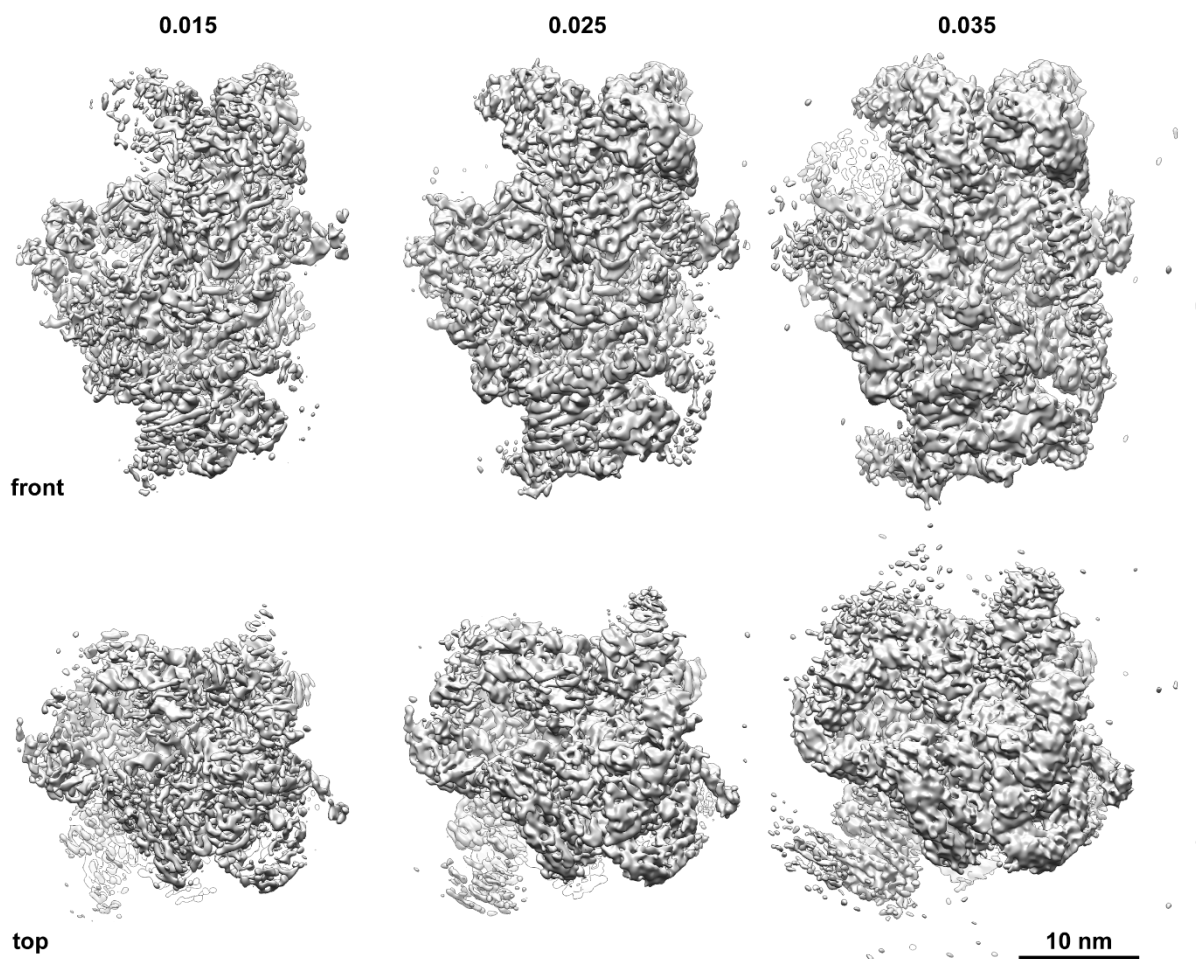


Figure 27. Depicts the *ct90S* map at different contour levels revealing the flexible regions in the periphery of the particle. From left to right the absolute values are 0.035, 0.025, and 0.015. The second row shows top views of the densities.

There is a considerable degree of flexibility in the periphery of the particle, coinciding with the resolution drop. There are two regions where considerable density appears at a lower contour level. One such region is located between foot and the body. The second low resolution cloud is the extension from a rod-like tether descending in a spiral from the head towards the left-side of the body. Along this path the low resolution region expands considerably in the back of the body and could encompass a larger sub-complex.

Segregation

The first step in obtaining a better understanding of the density was to segregate the clearly distinguishable dsRNA helices from proteins and single stranded RNA (ssRNA). This was achieved by placing markers by hand into the density and applying a radius for segregation. Segregating the map with Segger did not work well, since it is constrained to a particular contour level and the degree of flexibility in the map was too high for a meaningful segregation (Pintilie et al., 2009). The manually segregated map showed that the dsRNA is mainly concentrated at the core with radial protruding helices towards the periphery in the body and foot region (Figure 28). What became apparent after dsRNA segregation was that dsRNA in the head of the particle is concentrated at the core forming a cone towards the body. On the far left side of the body a long dsRNA helix was observed and on the opposite side a T-shaped RNA element was standing isolated from the rest of the body, which was used for the anchor point of the typical front view of the 90S. Taking a closer look at the long dsRNA helix showed it was sequestered by three ring-densities with discernible 7 blades typical of WD-repeat containing proteins. The T-shaped RNA protruded from the periphery towards the body where several dsRNA junctions were visible. At the foot of the particle three prominent dsRNA helices all pointed in the same direction towards the front, apart from that there were several single stranded rods in that region, but no complex junctions or similar present. The given resolution made it difficult to discriminate whether these small rods were protein helices or ssRNA.

The RNA core in the head region appears capped or surrounded by long solenoid densities composed in the cross-section of short rods perpendicular to the direction of the main tube. This was the first indication the pre-18S rRNA, which was thought to be located in that area based on the sc90S map, would be difficult to accommodate in an upright orientation. Such an orientation would position the head of the pre-18S in a region where there was considerably less dsRNA density found after segregation and it would be difficult to account for all the missing RNA especially since these capping solenoids resembled that of HEAT repeat domains. In addition, based on previous work on the characterization of the 90S particle components from other groups there were

several proteins in the 90S complex with predicted HEAT repeats. A comprehensive table describing individual protein components and summarizing their functions can be found in Woolford & Baserga, 2013. Based on the sizes of the HEAT repeat domain containing proteins the solenoids could be very long, which was in accordance with the observed densities in the *ct90S* map. Initially all the long solenoid densities in the head were thought to contain HEAT repeats, but as the resolution of the map increased the long solenoid densities started to become better resolved and the sub-structure appeared to be in some regions a mix of HEAT repeats and more flat densities also in an ordered arrangement. There were additional HEAT repeats observed in the map. An extremely well resolved repeat of α -helices was found at the back of the body and allowed for model building. The second prominent long alpha solenoid connects the foot of the particle with the head.

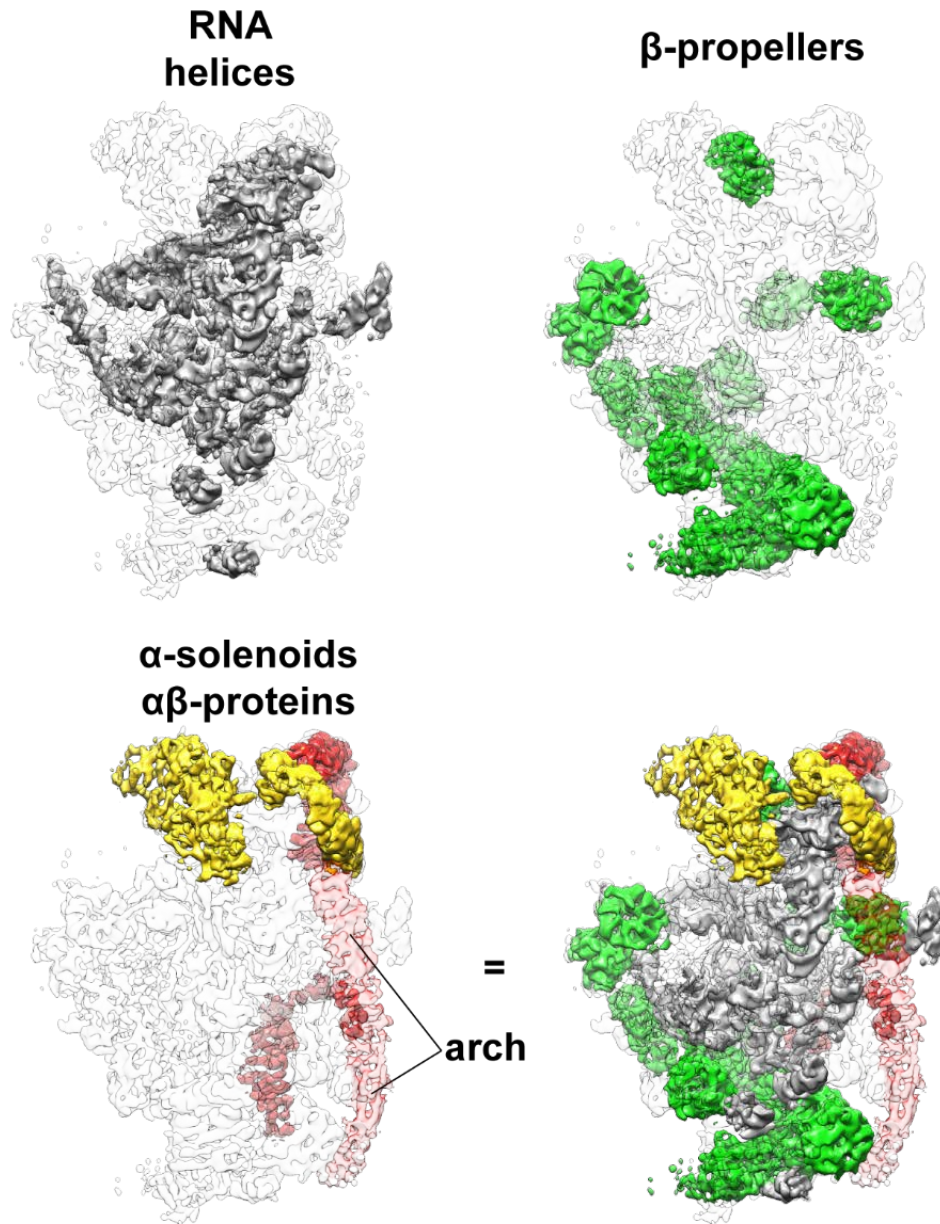


Figure 28. Density segregation of the *c*90S pre-ribosome. The dsRNA (grey) is concentrated at the core radially protruding towards the periphery. The β -propellers (green) are found mainly at the periphery of the particle building a shell. The α -solenoid resembling densities (red and yellow) are capping the head of the particle and connecting it with the foot. The elements containing in addition to α -helical repeats also β -strand repeats in the head are colored yellow. The bottom right density shows the sum of the segregated parts filling up much of the *c*90S density.

UTP-A and UTP-B

Located mainly at the periphery we could identify 19 β -propellers organized in two larger clusters of 9 and 6 β -propellers and 4 single β -propellers scattered throughout the map. From the protein predictions and literature two sub-complexes of the 90S the UTP-A and UTP-B were predicted to be rich in β -propellers. UTP-B in yeast has a molecular weight of 525 kDa and is comprised of four double β -propellers Utp1/Pwp2, Utp12, Utp13, and Utp21, one single β -propeller Utp18 and a predicted HAT-repeat containing Utp6 (Dosil and Bustelo, 2004). Furthermore, crosslinking data indicate the double β -propellers interact in pairs. One pair is made of Utp12 and Utp13 and the second pair is formed by Utp1 and Utp21 (Yang et al., 2012a). This same study showed also that the single β -propeller Utp18 crosslinks only to the second pair. Importantly, the crosslinks between both pairs are contained in their C-terminal regions only. Indeed we found a cluster of four double β -propellers and a nearby single β -propeller in the backpack part of the 90S body.

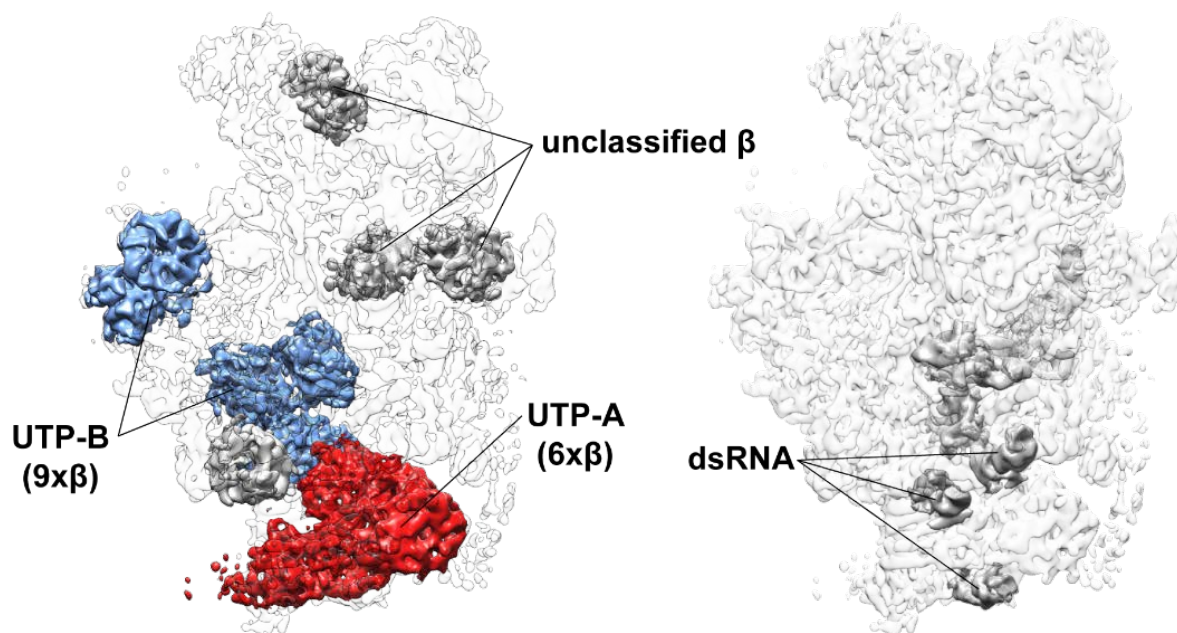


Figure 29. Left: The β -propellers assigned to sub-complexes UTP-B (blue) and UTP-A (red). The four single β -propellers (gray) are Sof1, Enp2, Utp7, and Rrp9. Right: the proposed 5' ETS of the α 90S. The densities for the pre-18S and the U3 snoRNA were subtracted from the dsRNA found.

Another observation was an α -helical bundle situated between three pairs of double beta propeller domains in the backpack region of the 90S. From the solvent side a clear four-fold symmetry was observed. The α -helical density was initially well resolved, but individual α -helices could not be clearly distinguished from each other. To improve the resolution the data were re-processed until the α -helices became distinguishable. A second part used for the map resolvability benchmarks was an α -helical bundle, which was later found to be the oligomerization domain formed by Nop56 and Nop58. The few days old improved map at 7.3 Å was given to Dr. Jingdong Cheng with the specific task of improving the given 18S 5' domain and central domain fitting, positioning dsRNA in segregated regions and tracing the ssRNA where possible.

The improved map allowed fitting of four bundles of α -helices (three short helices and one longer). Since these bundles interact via their long α helices in pairwise fashion of two and two this homo-tetrameric bundle is thus a dimer of dimers. PHYRE2 (Kelley et al., 2015) fold predictions for Pwp2, Utp21, Utp12, Utp13 found an α -helical C terminus for these proteins. It was previously reported Utp1/Pwp2 contains at its C terminus a coiled-coil domain whose function is unknown (Dosil and Bustelo, 2004). In accordance with the crosslinking data the oligomerization cluster what we observe in the density are the C-termini of Utp12, Utp13, Utp1, and Utp21 (Figure 30).

Next step was the elucidation of the identity of these β -propellers. For yeast Utp21 a crystal structure exists (Zhang et al., 2014). Assuming a yeast Utp21 structure would approximately enough resemble the fungus homolog I tried rigid-body fitting. Despite the characteristic open clam-shell orientation of the β -propellers two possibilities existed. Utp21 could be equally well fitted in two of the double β -propellers on opposite sides of the oligomerization cluster in the z-axis. An additional constraint was used to determine the identity of Utp21. Utp6 is the only UTP-B protein that does not contain a β -propeller domain, but is instead predicted to contain a HAT (Half-A-TPR) domain that interacts with the N-terminus of Utp21 (Champion et al., 2008). In the α 90S map there is only one such possible fit of Utp21 next to a solenoid density. In addition, the fitted Utp21 is oriented towards this solenoid with its N-terminus. Highlighting the interacting peptide sequence between residues 267 and 279 (GTSSGDLIFYDLD) shows in close proximity

a density touching the adjacent solenoid. From the density of the solenoid it is not clear whether this is a HAT or a HEAT repeat, what is clear though is that the solenoid density could harbor a protein the size of the Utp6 and not only a short region as predicted.

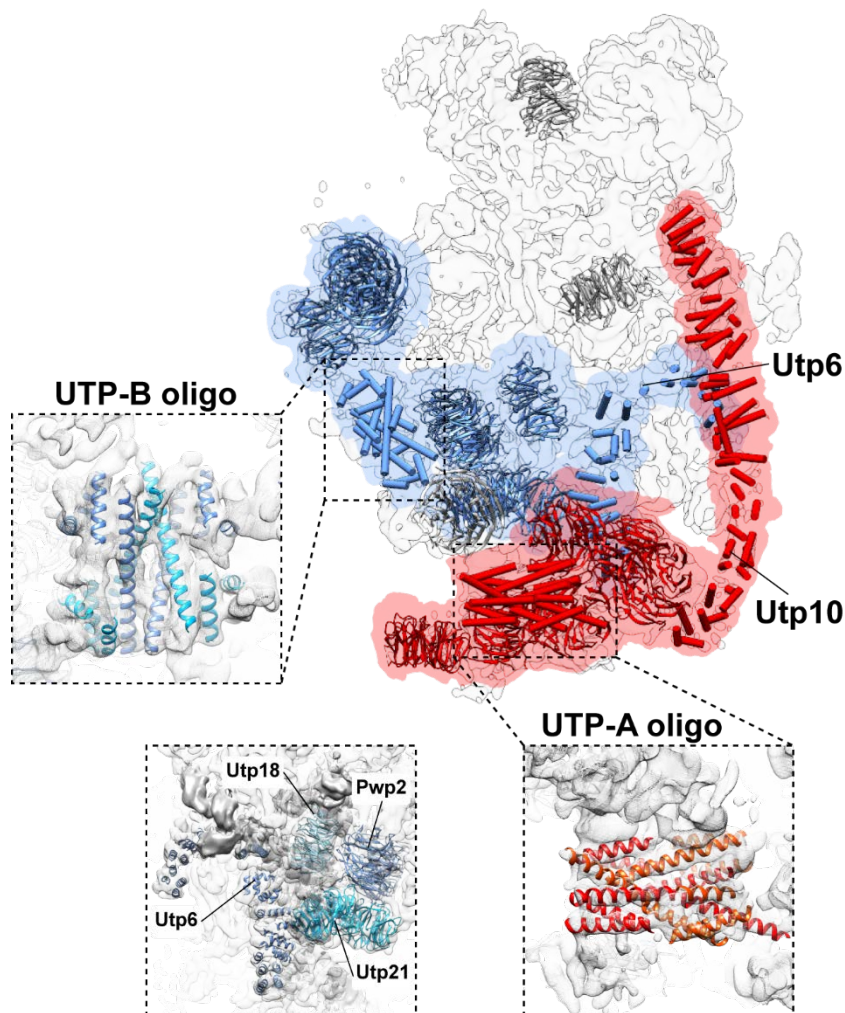


Figure 30. Oligomerization domains of UTP-B and UTP-A and the zoom into the Utp21 interaction to Utp6 and assignment of Pwp2/Utp1 and Utp18.

The second cluster of β -propellers is located at the foot of the ct90S pre-ribosome and interacts with three distinct dsRNA helices. This cluster coincides with the predictions for UTP-A sub-complex proteins harboring six β -propellers (Utp4 and Utp17 contain

double β -propellers and Utp8, Utp15 single β -propellers). Since this is the only such cluster in the structure left UTP-A was assigned to the foot of the α 90S. Supporting this assignment are the three dsRNA helices the β -propellers interact with. UTP-A was shown to bind to the 5' end of the 5' ETS sequence first, before other sub-complexes are recruited. The 5' ETS secondary structure in yeast predicts dsRNA helices at the 5' end of the 5' ETS with appropriate sizes to the helices found in the foot of the particle. Since the resolution was limited a clear connectivity between the helices could not be established and a clear 5' end was not found directly from the density.

Following the analogy of an oligomerization domain found in UTP-B a similar alpha helical bundle was found in UTP-A by Dr. Jingdong Cheng. The oligomerization domain of UTP-A is not as well resolved as in the UTP-B sub-complex. In contrast to the dimer of dimer in UTP-B, what can be observed differently in UTP-A is that the cluster does not follow a four-fold symmetry, but is comprised of two heterogeneous dimer pairs. The resolution was limited in that region and it was not possible to resolve to which proteins this oligomerization elements connect with.

U3 snoRNP

The U3 snoRNP plays a crucial role in coordinating the early cleavages A_0 , A_1 , and A_2 in the 90S particle (Hughes and Ares, 1991). The crystal structures of the U3 snoRNP constitutive proteins Nop56, Nop58, Nop1 and Snu13 could be unambiguously fitted in the body of the 90S particle between the Utp10 and the Utp6 by Jingdong Cheng. Nop56 and Nop58 form a clearly visible oligomerization domain with their long α -helices. The dsRNA density in proximity to the fitted proteins was therefore assigned to the U3 snoRNA. The density allowed for building a model of the U3 snoRNA from its 3' hinge region to the 3' end based on the secondary structure information from yeast. The three-way junction of helix 3 and the T-shaped element of helix 4 provided two starting points for model building (Figure 31). Additional constraints were derived from previous crosslinking data of the U3 snoRNP proteins (Granneman et al., 2009). Rrp9 a WD40 repeat domain protein associated with U3 snoRNP was fitted next to the T-shaped dsRNA corresponding to parts of helix 2 and 4 of the U3 snoRNA model. Rrp9 was previously shown to crosslink to parts of helix 2 and 4 (Granneman et al., 2009) and

since there was only one single β -propeller in that region the Rrp9 assignment was unambiguous despite the density for this particular β -propeller was not extremely well resolved. We can also trace in the density the heteroduplex formed by the 3' hinge region of U3 snoRNA and the 5' ETS helices V and VI pointing towards Utp1/Pwp2 and Utp21 of the UTP-B complex.

The overall well resolved density for this area suggests U3 snoRNP is rigidly anchored at the side of the 90S body reaching towards the core and its U3 snoRNA appears to be stabilized by the protein interactions. A 70 Å long α -helix in proximity to the heteroduplex contacts the U3 snoRNA. In the density we observe two even longer α -helices forming a letter V that engulf the RNA-rich core of the 90S from both sides and at the apex bind a long dsRNA. The strategic positions of these three α -helices suggest a possible role in the disassembly of the 90S complex.

Unfortunately the region from the 3' hinge towards the 5' end of U3 snoRNA which is involved in the hybridization with the 18S rRNA and therefore cannot form the central pseudoknot at this stage is buried in the core of the particle. Despite the well resolved density in that region the close packing of proteins and RNA prohibits an unambiguous interpretation of the density at the present resolution.

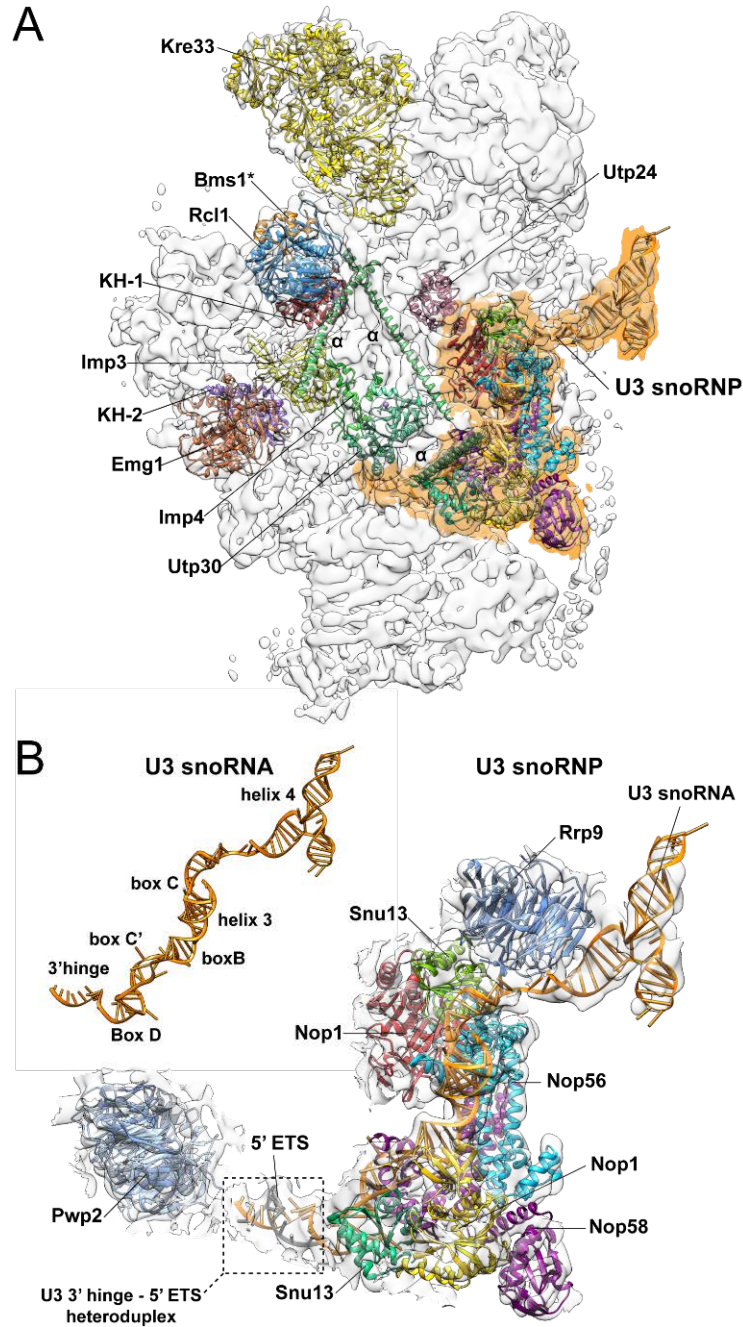


Figure31. A: The single proteins fitted that are not classified as part of a sub-complex in the 90S. Three long α -helices (green) are found at the center of the particle. B: A view into the U3 snoRNP. The U3 proteins are from *S. solfataricus* and in top left a model of the built U3 snoRNA (orange) is shown. The hybridization between the 3' hinge of U3

snoRNA and the complementart 5' ETS sequence (gray) is in proximity to the Pwp2/Utp1 double β -propeller.

18S

The 18S rRNA is oriented upside down relative to the canonical 40S view with the 18S head facing up and the inter-subunit space on the right. We could identify most of the 5' domain, the platform of the central domain and only a small part of the 3' major domain (Figure 32).

18S – 5' Domain

The 5' domain spans between the head and body region and resembles to a large extent the mature 18S. Since there is no *C. thermophilum* ribosome crystal structure available we fitted the *S. cerevisiae* 18S (Ben-Shem et al., 2010). The mature yeast 5' domain rRNA is fitting extremely well into the α 90S density. With the exception of a few helices we could find most of the 5' domain helices were already folded and located at their canonical sites. For example, we observed the hairpin loop of h11 interacts with h7 resembling the mature state. Similarly the major minor packing of h6 and h8 resembles its canonical state. The co-axial packing of h16 and h17 known as the shoulder is also present, however relative to the rigid-body fitted 5' domain helices h7-h11 and h6-h8 this pair of helices is slightly shifted away from its mature position, suggesting a further compaction still needs to occur. The hairpin loop of h16 is rotated or skewed relative to the mature h16 orientation. However, this could also be a species specific feature. Several ribosomal proteins of the 5' domain could be identified at their mature positions: eS4, uS4, eS6, eS8, eS17, and eS24. The density for the N-terminus of eS24 is missing in its canonical binding position. Surprisingly, already at this stage the eS6 containing a long α -helix is positioned in the mature state.

18S – Central Domain

Following the direction from the last 5' domain helix we encountered a part of the central domain at the back of the body. We identified h20, h22, h23 and h26 which constitute the platform of the 18S. This bundle of helices resembles the maturely folded 18S, with the exception of h24a fragment missing from the density. Having aligned the mature

18S to the 5' domain, the position of the pre-platform domain is slightly shifted compared to the mature position. We could not unambiguously identify the four extension segments of ES6. There is plenty of unallocated density resembling dsRNA at the back of the body. Parts of the crystal structure of the extension segments could be rigid-body fitted, but not the complete mature extensions, suggesting this part of the 18S is in an alternate conformation. Unfortunately the density in that region is not resolved well enough to enable tracing of ssRNA and the extension segments are not structurally differing in size to a degree at which one could clearly assign their individual identities. Therefore, we could not fit parts of the extension segments reliably into the density that would provide a unique fitting solution.

As observed for the 5' domain, the proteins eS1, eS7, uS8, uS11 and uS15 interacting with the platform were found at position corresponding to their canonical positions in the mature 18S. One should note the density is missing for the N-terminus of uS15. For the small ribosomal protein uS11 we have density at the canonical position, but the protein itself is small and does not contain secondary structure features making therefore the identification difficult relying on structural constraints only. We made the assignment in part relying on the structural information, but also on the basis of the strong mass-spec signal for this protein. The assumption was that the likelihood that a different protein of similar size would occupy this exact position and at the same time uS11 would bind somewhere else in the 90S particle to later move to its canonical binding site, would be rather low.

18S – 3' Major Domain

The core of the *ct*90S particles is well resolved, however RNA and proteins are densely packed in this region which makes it difficult to clearly discern between individual components. We could identify one cluster of the 3' major domain helices h28, h29, h30, h41 and h43, with their canonical ribosomal binding partners uS7, uS8 and eS28 in place. This cluster will later form the back part of the 40S head. Readily discernible is the density for the 65 kDa protein Emg1 located near uS7 and eS28, which positions it near its binding sequences in the 18S rRNA.

Other parts of the 3' major domain could not be unambiguously identified in the density. The smaller the segment sizes of the rRNA we tried to search for in the cryo-EM map, the less reliable such a constraint has become, since not only one position would then satisfy the fit anymore. A task we found particularly difficult in this densely packed region with a high level of interwoven proteins and RNA. The lack of recognizable larger segments of the mature 3' domain including the characteristic beak, are an indication that this domain still needs to undergo considerable folding before its mature state can be reached.

18S – 3' Minor Domain

The 3' minor domain mainly consisting of a characteristic long h44 that is involved in mRNA decoding in the mature 40S could not be unambiguously identified. There is no density for a long dsRNA near the canonical binding position of h44. However, adjacent to the fitted parts of the central domain is a long dsRNA helix of appropriate size, which is held in check by three UTP-B complex β -propellers and reaching towards the 3' minor domain, but at this stage of assembly the h44 is expected to be hybridized with parts of ITS1 that inhibit the pre-mature activation of the decoding site and therefore this might result in structural changes prohibiting rigid-body fitting of the mature h44 (Lamanna and Karbstein, 2011).

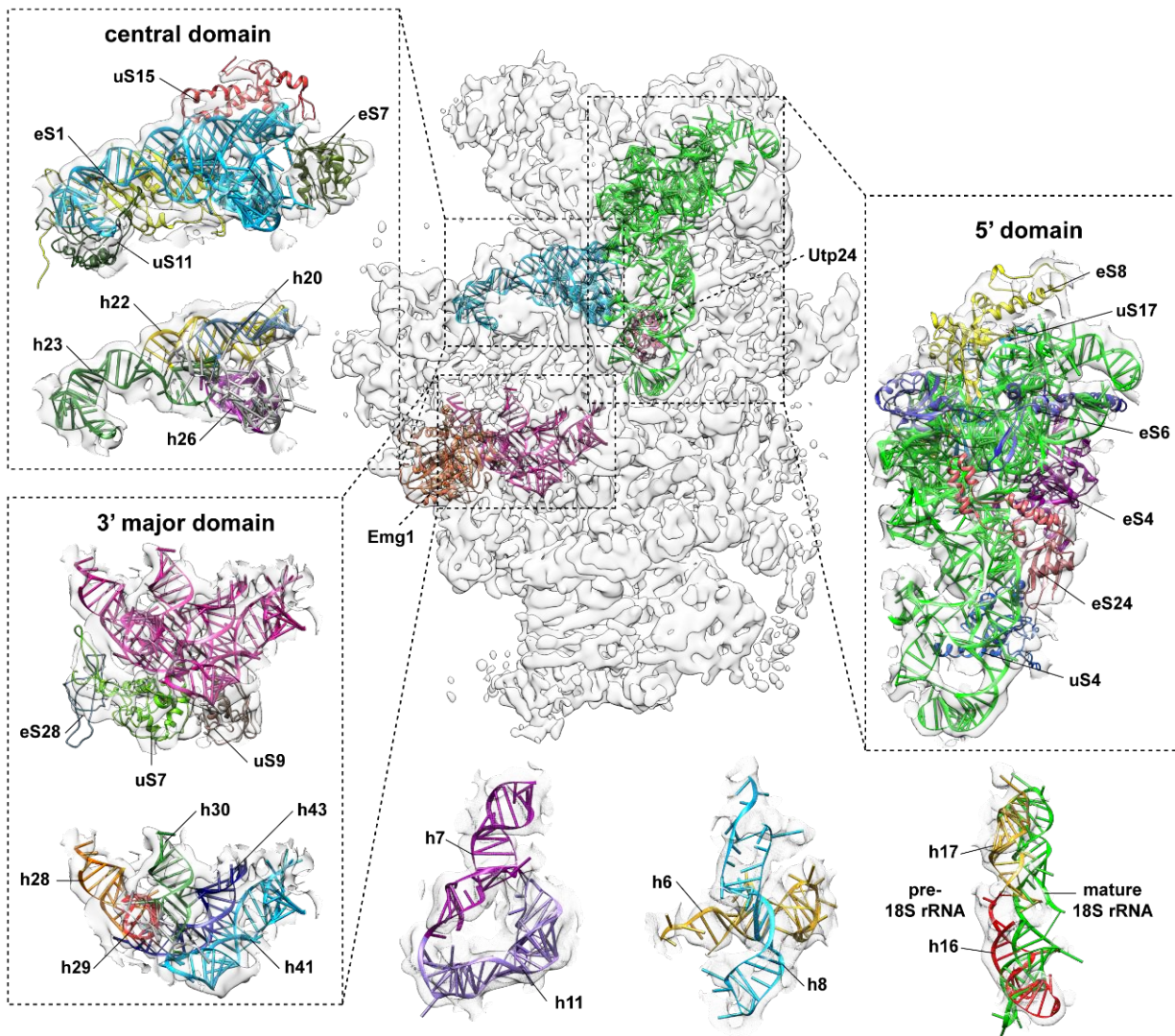


Figure 32. The fitted pre-18S rRNA in the $c90S$ density. The r-proteins and the RNA helices are shown for each fitted 18S domain (5' domain, central domain, 3' major domain). Bottom row: the 5' domain helices resembling their mature fold and the shifted orientation of the h16/h17 stack compared to the mature 18S rRNA.

18S – Summary

Comparing the orientations of the three fitted domains; the 5' domain, central domain, and the 3' major domain to the domains of the mature 18S it becomes apparent that as we progress from the 5' end to the 3' end of the 18S the state of maturity of the

individual domains decreases. For the 5' domain most of the RNA and proteins were present and in the mature state lacking only a final compaction step. The central domain platform was found assembled albeit shifted with respect to its mature position. The extension segments are not found located at their mature positions. In addition, the base-pairing interaction of ES6d with the 5' domain rRNA is not present at this stage. The 3' major domain is least resembling the mature state and what is even more striking is that the part we could identify is located ~10 nm away from the central domain platform. In addition to the substantial shift the 3' major cluster has to extensively rotate in subsequent maturation steps and assemble with the other parts of the 3' major domain (Figure 33). Therefore, before reaching its mature state the central domain has to undergo vast structural changes.

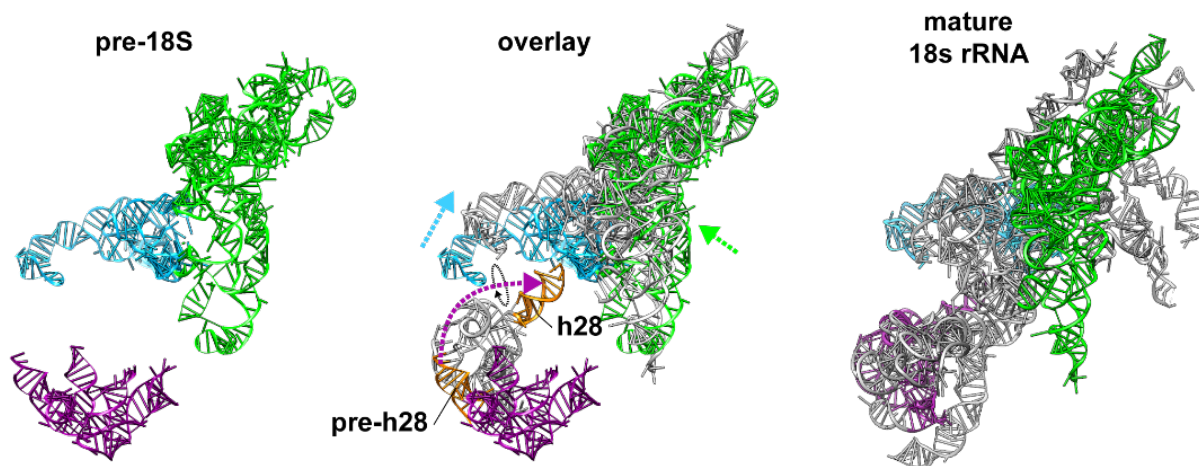


Figure 33. The comparison of the pre-18S to the mature 18S shows the 5' domain is slightly shifted, the central domain is shifted more, and the 3' major domain has to undergo vast structural rearrangements to reach the mature orientation as shown in the overlay (middle).

Discussion

90S and Terminal Knobs

The 90S pre-ribosome is a large ribonucleoprotein complex comprised of a number of protein factors, r-proteins and more than 2,5 kbases of RNA. These components are arranged in an intricate network of RNA and protein contacts forming a large and defined complex. This is very much in contrast to the observations derived from the Miller spreads where terminal knobs were indeed observed, but the rRNA branches ending in these knobs were linearly spread out and not compacted as seen in the 90S pre-ribosome. Terminal balls were observed across all eukaryotes (Mougey et al., 1993), and it was believed the high degree of conservation across species offers an evolutionary advantage. However, the spreading conditions did not entirely preserve the structure and therefore the formation of a large compact particle was missed for decades. Modifying the Miller spreading technique to more physiological conditions led to the observation of larger terminal knobs preserving more of the RNA-protein contacts (Osheim et al., 2004). These large knobs measured 35 to 45 nm in size, which is similar to the 36 by 40 nm we observe for the 90S pre-ribosome. One has to note that the original Miller spreads were a remarkable achievement and offered the first view in the organization of active rDNA genes and are still being found in many modern textbooks, which is a rare feature in a rapidly developing field such as this (McKnight et al., 2012).

Is 90S Necessary?

Finally that we are able to visualize the elusive 90S complex a simple question arises. Is the formation of a large particle dedicated to the maturation of the small (40S) subunit really necessary? It seems bacteria can make self-assembling ribosomes where the information sufficient to form a ribosome is contained in the individual components. For example a fully functional *E. coli* small (30S) subunit was successfully reconstituted by simply adding the purified components together (Traub and Nomura, 1968). However, the efficiency of the 30S reconstitution is extremely low at around 18% when the 21 r-proteins and the 16S are added simultaneously together (Culver and Noller, 1999). It

was shown that the r-proteins bind to the 16S in a defined sequence associating with the 5' domain first and the 3' end last (Mizushima and Nomura, 1970). When considering the hierarchy of assembly and adding r-proteins in a step-wise manner the yield considerably increases to 45% (Culver and Noller, 1999). Even *in vivo* the 30S subunit assembly sometimes leads to dead-end products. The temporal assembly order in which r-proteins bind to the 16S is extremely important and the lack of a more controlled environment results in parallel assembly pathways not always leading to functional ribosomes (Mulder et al., 2010). However, *E. coli* has approximately 60 assembly factors, majority of which are not essential, but they are involved in increasing the efficiency of the process and to prevent premature subunit engagement (Stokes et al., 2014).

Efficiency

It would be interesting to compare the total amount of rRNA transcription to the total number of produced ribosomes to get an approximate view into the efficiency of the process in eukaryotes. Yeast is known to contain approximately 150 rDNA repeats, half of which are active in the exponential growth phase. The average speed of the RNA polymerase I has been approximated to be around 60 or 40 nucleotides per second (French, Osheim, Cioci, Nomura, & Beyer, 2003; Kos & Tollervey, 2010). The time it takes for one RNA polymerase to transcribe the 6600 nucleotides long rDNA can thus be approximated to somewhere between 110 and 165 s. The tracing of the Miller spreads showed several RNA polymerases engage in transcription of an active rDNA locus, similar as multiple ribosomes can simultaneously translate one mRNA. From about 60 RNA polymerases transcribing one active rDNA at any given time (Figure 5) and each polymerase requiring roughly 2 min to make one primary transcript, this gives $30 \times 75 = 2250$ rRNAs/min, which is about 110% of the final 2000 ribosomes produced every minute (Warner, 1999). A slightly higher number can be estimated from a transcription counting study where it was shown that every 1.6 s a new RNA polymerase I initiates in transcription (Tan et al., 2010). Thus every minute $37 \times 75 = 2775$ full rRNAs are produced resulting in a 40% error margin. These crude estimates are to be taken with a grain of salt, since the comparison is taking into account the total

production of the small and large ribosomal subunits and not the initial step related to the 90S pre-ribosome. However, the overall efficiency of the assembly does suggest that the ribosome biogenesis in yeast has a relatively low failure rate.

Energetics

Building a scaffold to ensure proper folding, processing and modifications of the nascent pre-18S rRNA is a huge investment for the cell. It seems counter-intuitive that such an effort can be justified in energy terms. If such a complex significantly reduces the occurrence of trapped intermediates and in a way helps increase the rate of productive outcomes, then this might balance out the initial energy investment. In addition, the scaffold does get recycled and the components can be reused after the dissociation of the 90S complex. However, is energy really the most important factor to consider? A high fidelity ribosome assembly is producing fewer defect pre-ribosomes that have to be degraded and thereby the stress on the nuclear surveillance is reduced. Where the fitness and survival of individual cells is key, quality is more important than quantity. If one thinks of bacteria as optimized for speed to rapidly outgrow their competitors, the eukaryotes might have gone down a different path in evolution. Preferring quality and control over speed might have been an evolutionary advantage leading eventually to multi-cellular organisms. At any rate the 76 snoRNPs and over 200 associated factors in yeast (over 280 in humans) is still surprisingly high (Woolford, 2015). Part of the factors can be attributed to the transportation in and out of the nucleus. However, the significant advantage a structure such as the 90S complex offers is regulation and control over the individual steps of assembly best illustrated from a structural point of view.

Structure Defines Function

The 90S pre-ribosome is a great example of how structure defines function at the level of macromolecules. The 90S complex formation starts immediately after transcription begins. The 5' ETS sequence provides the binding string to which sub-complexes UTP-A, UTP-B, U3 snoRNP and other factors associate and form a scaffold that will accommodate the pre-18S rRNA once the RNA polymerase I has progressed far enough along the rDNA. The partially folded pre-18S rRNA is attached or pulled to the

scaffold by long tethers, capping the 5' domain and connecting the head of the 90S particle with UTP-A located at the foot (Figure 34). The central domain and the 3' major domain both interact with the UTP-B and U3 sub-complexes at the core of the particle. The many β -propellers of UTP-B provide a solid structural foundation to which several RNA helices are attached. The RNA-protein contacts are responsible for an open structure with many cavities between the inner core and outer shell providing ample space for modifying snoRNPs to access the many sites of methylation and pseudo-uridylation on the rRNA. At the same time processing of the spacers is guided by the U3 snoRNP, which also prevents pre-mature 18S compaction by base-pairing to specific regions within the pre-18S. The open structure provides a controlled environment allowing local folding and modifications to have sufficient time to take place in and the r-proteins to incorporate. After the dissociation of the 90S complex the released pre-40S is ready for compaction and the 5' ETS particle undergoes recycling. In simple terms, the 90S complex divides the complex 18S folding into smaller steps and provides time for the maturation of the nascent small ribosomal subunit.

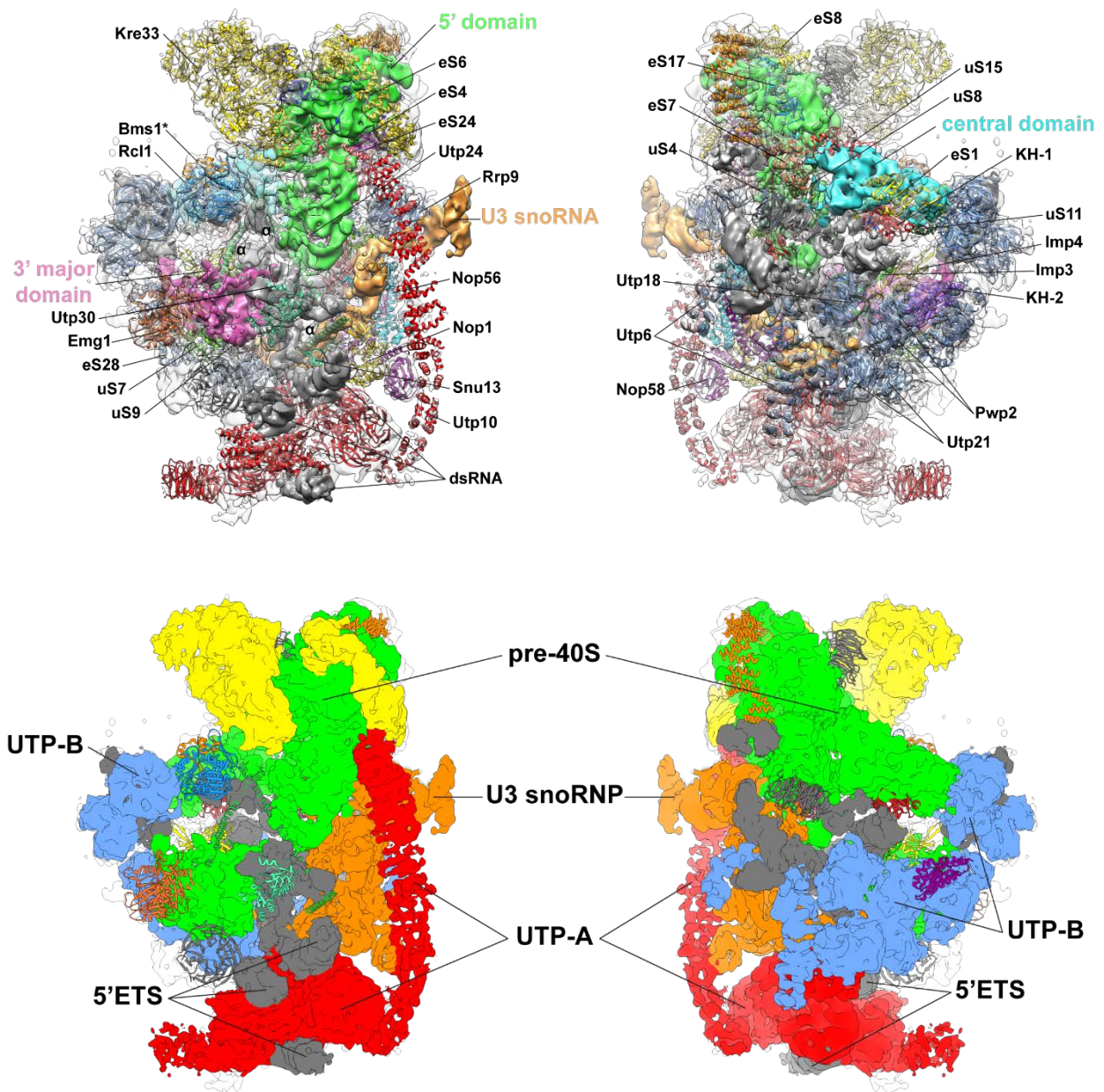


Figure34: Top: A structural model of all fitted protein crystal structures and segregated RNA as densities. Bottom: A cartoon representation showing the sub-complexes and the pre-40S containing moiety.

β -propellers and α -solenoids

Utilizing β -propeller folds and α -solenoids is a common pattern in the 90S complex, where the β -propellers are providing a rigid support for the scaffold and long range connections are α -solenoid based. On the protein level a prominent example is found in the UTP-B sub-complex. UTP-B has four proteins containing N-terminal double β -propellers and C-terminal α -helical domains. The same order of WD40 repeats at the N-terminal regions followed by α -solenoids at the C-termini was observed for the vesicle coat complexes and the nuclear pore complex (NPC) (Devos et al., 2004), and for some proteins in the intraflagellar transport (IFT) (van Dam et al., 2013). These β -propeller folds and α -solenoid like motifs enable membrane curving, which is a hallmark feature of eukaryotes. However, there is no known interaction with membranes in the 90S complex, suggesting rather a structural role of these components. Why the observed order beta-followed-by-alpha is evolutionary conserved remains a mystery.

U3 snoRNP

We were able to locate the U3 snoRNP sub-complex based on the crystal structure of the butterfly resembling active C/D box RNP (Lin et al., 2011). The U3 has not been shown to be involved in the methyl group transfer typical for C/D box RNPs, and compared to the active C/D box RNP structure we accordingly find the methyltransferase Nop1 in an alternate location. Instead of guiding modifications the U3 snoRNP is involved in processing the pre-rRNA. We find the U3 reaching towards the RNA-rich core of the 90S where it is thought to base-pair with complementary regions close to the early cleavages A₀, A₁ in the 5' ETS and A₂ in the ITS1 (Dutca et al., 2011) and in addition to complementary sites in the pre-18S rRNA preventing pre-mature pseudoknot formation (Kudla et al., 2011). The formation of the central pseudoknot is the key step in spatially coordinating the 5' domain, central domain, and 3' major domains, and is required to occur before the small subunit can engage in translation. Thus the U3 can act as a molecular timer, allowing after the 90S dissociation event the collapse of the 18S domains into a more mature pre-40S particle. The trigger responsible for the removal of the U3 snoRNP was recently identified as a DEAH-box helicase Dhr1 (Sardana et al., 2015). It would be interesting to test whether this helicase

alone is sufficient for causing dissociation of purified 90S pre-ribosomes or if additional factors are needed.

18S Folded Stages

The pre-18S found in the 90S complex is already partially folded. The 5' domain is to a large extent resembling the mature state. There are small differences found in several hairpin loops, but it is difficult to say these differences are a result of an immature folding state or species related, since we used a *Saccharomyces cerevisiae* crystal structure (Ben-Shem et al., 2010) as a starting model for the *Chaetomium thermophilum* 90S fitting. The central domain is found less resembling the mature state. We find the platform folded, although slightly shifted with respect to the mature 18S. The extension segments were not unambiguously identified. There are dsRNA helices in that region, but additional constraints would be required to allow fitting of individual parts of the extension segments. The 3' major domain is least resembling the mature state. The gradual decrease in 18S maturation from the 5' to the 3' end is similar to what was observed for the 16S in bacteria (Mulder et al., 2010; Sashital et al., 2014). Not surprising the r-proteins follow a similar hierarchy of assembly coinciding with the pre-18S gradual maturation as was found for bacteria and yeast (Jakob et al., 2012; Sashital et al., 2014) The concept of sequential and hierarchical assembly dates back to the Nomura assembly map (Nomura, 1973). The r-proteins are thought to stabilize secondary RNA structure interactions by predominantly binding to the surface of the rRNA and in doing so inducing conformational changes in the rRNA structure allowing for a second set of r-proteins to associate.

The 90S structure seems to support maturation of the 5' domain, but not the 3' end. The 3' minor domain with the typical h44 which contains the decoding site is at this stage hybridized to ITS1 to prevent premature decoding site exposure (Lamanna and Karbstein, 2011). The ITS1 is not removed from the pre-18S before the pre-40S reaches the cytosol. Since a RNA sequence with a strong helix formation propensity would be expected to fold into its secondary structure spontaneously, it would be convenient to speculate the h44 is already structured, forming an ITS1 hybrid at this stage. There is a long dsRNA the size of h44 sandwiched between three β -propellers of

UTP-B. In order to reach the mature position of h44 all that would be required for this particular helix is a clapping movement towards the 5' domain after the UTP-B complex blocking the way has first dissociated.

Utp10

Another interesting protein is Utp10, of which we find an approximately 250 Å long alpha solenoid reaching in an arch towards the head region. To build this long arch roughly 1100 residues were needed whereas Utp10 is around 1800 residues long. From the 90S structure the polarity of the Utp10 is not clear. Utp10 could be traversing the UTP-A complex and touching the head region with either its C- or N-terminus. The Utp10 N-terminus was shown to form a complex with another UTP-A protein Utp17 (Kornprobst, Turk et al., 2016). From the perspective of assembly it is interesting to discuss its function. UTP-A is the first sub-complex to associate with the rRNA. What is interesting is that the region it interacts with is in the chronological order of assembly appearing after UTP-B, U3 snoRNP all have been recruited to the growing complex (Chaker-Margot et al., 2015). The question arises why the interaction with the head region is needed. More specifically Utp10 contacts an alpha beta solenoid, most likely Rrp5, in vicinity of the 5' domain of the 90S head. For now it remains a mystery why such a long arch has been utilized and not for example the U3 snoRNP which is located closer nearby. Since Rrp5 is required for the UTP-C sub-complex association to the 90S (Pérez-Fernández et al., 2007), Utp10 might be the initiator of a long assembly cascade reaching over the complex towards the opposite side of the 90S structure, where UTP-C is most likely located in a low resolution region, based on the crosslinking constraint between the Rrp7 protein of UTP-C with h26 of the central domain (Lin et al., 2013).

α-Helices

Puzzling are also the roles of the three long α-helices found in the structure. Two of which hold the RNA core from both sides and the third helix, slightly shorter, is in proximity to the U3 snoRNP. The release of the particle is mediated by a helicase (Dhr1), but is the unwinding of the RNA enough for the dissociation of the 90S complex? It is tempting to speculate given the strategic position of these three helices,

that if one would exert a pulling force on these helices the particle would be able to collapse, at least in the central region.

Kre33

Kre33 acetylates two cytosines important for translation accuracy in h34 and near the decoding site in h45. What is so unusual about this protein is where we find it in the 90S complex. It makes little sense why a modifying enzyme would be located more than 10 nm away from its target sequence. In addition, Kre33 associates with the appearance of the 3' minor domain and binds the 5' domain in the 90S head which has been sitting there for a while, suggesting additional factors coinciding with the 3' minor domain recruit Kre33 to that region. Kre33 is located where one would find in the mature 18S part of the extension segment (ES6). The empty cavity we observe in the 90S pre-ribosome between the platform and the Kre33 will be occupied in the mature subunit by the extension segments and h44. Therefore, in that region considerable rearrangements have to occur after the pre-40S is released from the 90S complex.

Integrative Modelling

The problem of finding the missing extension segments that could not be unambiguously assigned could be addressed using an integrative modeling approach (Russel et al., 2012). The approximate location of the extensions is known due to connectivity constraint linking the extensions to the major part of the central platform, which was identified in the density. There is a limited space of unassigned density in that region to search in. The 18S rRNA sequence is known as well as the secondary structure of the extension segments. Information regarding the density, connectivity, secondary structure can all be used as constraints. The hypothesis could be tested if the unassigned dsRNA densities we observe could accommodate the missing helices of the extension segments and more importantly whether or not the constraints at hand are sufficient to find a unique fitting solution. In such a framework additional constraint such as keeping the known nucleotide modification sites solvent accessible can be readily added. The more pieces of the puzzle are fitted and more information is available the more the system is constrained and the degrees of freedom reduced. One

difficulty is that it is not known to what extent the helices in question are folded in a mature state and not in an alternate conformation as the expected h44::ITS1 hybrid, but it would be interesting to try and find out.

Biogenesis Highway

Given the large numbers of ribosomes made every minute one asks the question whether or not a gradient of diffusing maturing pre-ribosomes exists from the nucleolus towards the NPC. If there is a highway one would expect to see a path consisting of many maturing subunits. If such a transportation highway might exist, its visualization should be feasible with tomography.

Stage Specific Constructs

To grasp the transitions states of the 90S maturation it would be interesting to design specifically staged intermediates, where each step would be assembled *in vitro*, visualized using single particle cryo-EM, and by adding the relevant components the intermediate would be allowed to progress to the next stage. The complexity of the process makes such an approach difficult in practice. The first step in that direction would be the reconstitution of the 5'ETS complex.

The 5' ETS Beginning

The wealth of information the *ct*90S offers was hopefully demonstrated. Now that we have the structure of the yeast and thermophile 90S complexes, comparisons can be made. Unfortunately the *sc*90S reconstruction was made in the time before access to a direct electron detector was available, limiting considerably its resolution, but overlaying both electron densities showed one important difference. A clearly visible single β -propeller in the *sc*90S map is absent from the *ct*90S density even at extremely low contour levels. From the biochemistry we know the yeast 90S UTP-A complex contains the protein Utp9, a predicted single β -propeller, that is not present in thermophile fungus. The Utp9 was shown to crosslink within the first 100 nucleotides of the 5' ETS sequence marking the beginning of the rRNA transcript (Kornprobst, Turk et al., 2016). A higher resolution structure of the yeast 90S would be extremely interesting to unravel the beginning of the 90S complex assembly.

Assembly as a Process

A final simplified model of the assembly stages is provided, showing the assembly in three stages (Figure 35). However this model is purely speculative, since we only have structural data on stage III. The point of the model is to illustrate the small subunit assembly as a process. In stage I the sub-complexes assemble around the 5' ETS transcript. In stage II more assembly factors associate with the growing particle and the pre-18S rRNA is gradually being incorporated from the 5' domain first towards the 3' end. After the full 18S has been transcribed it takes approximately 40 s for the maturation of the 90S particle to occur. Followed in stage III is the mature 90S particle, which is awaiting dissociation into the pre-40S and 5' ETS (sedimenting at 35S) moieties, the later will be degraded and recycled by the exosome machinery to engage in another round of small subunit assembly.

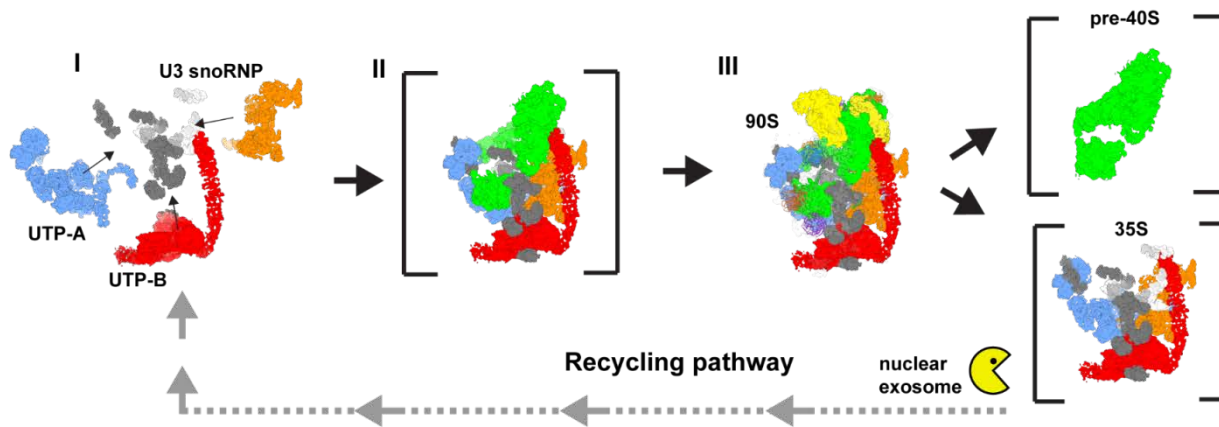


Figure 35: A simplified model of the 90S assembly.

Biogenesis as a New Target for Antibiotics

Early on when I started working on the project the idea emerged that the ribosomal biogenesis would be a terrific target to find new antibiotics. The differences in ribosome assembly between bacteria and humans are numerous. First both systems have to be understood to a point before biogenesis assembly steps can be specifically targeted, or in a brute-force approach a systematic screen using chemical compound libraries might lead to the discovery of biogenesis inhibitors. Recently there was to my knowledge a first small molecule discovered involved in inhibiting bacterial biogenesis (Stokes et al.,

2014). However, since bacterial ribosomes can self-assemble and the assembly factors are predominantly non-essential finding a biogenesis target might prove challenging. Ribosomal biogenesis might also be a potential drug target against protists, such as *Trypanosoma brucei* the causative agent of the sleeping disease. The current drugs on the market (with one exception) are from before 1950 and cause severe side effects. Since protists separated from our evolutionary branch early on it is likely that differences in the biogenesis of ribosomes exist, making them an interesting target for designing new antibiotics. In either case the potential of finding new mechanistic classes of antibiotics is there and I am surprised to find out the assembly of a ribosome has not been sufficiently appreciated in this regard. Perhaps now is a good time.

References

- A. Yonath, J. Muessig, B. Tesche, S. Lorenz, V.A.E. and H.G.W. (1980). Crystallization of the large ribosomal subunit from *B. stearothermophilus*.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science* (80-.). 289, 905–920.
- Beckmann, R., Bubeck, D., Grassucci, R., Penczek, P., Verschoor, A., Blobel, G., and Frank, J. (1997). Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. *Science* 278, 2123–2126.
- Beckmann, R., Spahn, C.M.T., Frank, J., and Blobel, G. (2001a). The active 80S ribosome-Sec61 complex. *Cold Spring Harb. Symp. Quant. Biol.* 66, 543–554.
- Beckmann, R., Spahn, C.M., Eswar, N., Helmers, J., Penczek, P.A., Sali, A., Frank, J., and Blobel, G. (2001b). Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell* 107, 361–372.
- Ben-Shem, A., Jenner, L., Yusupova, G., and Yusupov, M. (2010). Crystal structure of the eukaryotic ribosome. *Science* (80-.). 330, 1203–1209.
- Blobel, G., and Dobberstein, B. (1975). Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J. Cell Biol.* 67, 835–851.
- Brink, M.F., Verbeet, M.P., and de Boer, H.A. (1993). Formation of the central pseudoknot in 16S rRNA is essential for initiation of translation. *EMBO J.* 12, 3987–3996.
- Bunner, A.E., Trauger, S.A., Siuzdak, G., and Williamson, J.R. (2008). Quantitative ESI-TOF analysis of macromolecular assembly kinetics. *Anal. Chem.* 80, 9379–9386.
- Carazo, J.M., and Frank, J. (1988). Three-dimensional matching of macromolecular structures obtained from electron microscopy: an application to the 70S and 50S *E. coli* ribosomal particles. *Ultramicroscopy* 25, 13–22.
- Chaker-Margot, M., Hunziker, M., Barandun, J., Dill, B.D., and Klinge, S. (2015). Stage-specific assembly events of the 6-MDa small-subunit processome initiate eukaryotic ribosome biogenesis. *Nat. Struct. Mol. Biol.* 22, 920–923.
- Champion, E.A., Lane, B.H., Jackrel, M.E., Regan, L., and Baserga, S.J. (2008). A direct interaction between the Utp6 half-a-tetratricopeptide repeat domain and a specific peptide in Utp21 is essential for efficient pre-rRNA processing. *Mol. Cell. Biol.* 28, 6547–6556.
- Chen, J.Z., and Grigorieff, N. (2007). SIGNATURE: a single-particle selection system for molecular electron microscopy. *J. Struct. Biol.* 157, 168–173.
- Chimnaronk, S., Suzuki, T., Manita, T., Ikeuchi, Y., Yao, M., and Tanaka, I. (2009). RNA helicase module in an acetyltransferase that modifies a specific tRNA anticodon. *Embo J.* 28, 1362–1373.
- Conconi, A., Bessalov, V.A., and Smerdon, M.J. (2002). Transcription-coupled repair in RNA polymerase I-transcribed genes of yeast. *Proc. Natl. Acad. Sci. U. S. A.* 99, 649–654.

- Crick, F. (1988). *What mad pursuit : a personal view of scientific discovery* (New York: Basic Books).
- Culver, G.M., and Noller, H.F. (1999). Efficient reconstitution of functional *Escherichia coli* 30S ribosomal subunits from a complete set of recombinant small subunit ribosomal proteins. *RNA* 5, 832–843.
- van Dam, T.J.P., Townsend, M.J., Turk, M., Schlessinger, A., Sali, A., Field, M.C., and Huynen, M. a (2013). Evolution of modular intraflagellar transport from a coatomer-like progenitor. *Proc. Natl. Acad. Sci. U. S. A.* 110, 6943–6948.
- Delprato, A., Al Kadri, Y., Perebaskine, N., Monfoulet, C., Henry, Y., Henras, A., and Fribourg, S. (2014). Crucial Role of the Rcl1P-Bms1P Interaction for Yeast Pre-Ribosomal RNA Processing. *Nucleic Acids Res.* 42, 10161.
- Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B.T., Sali, A., and Rout, M.P. (2004). Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol.* 2.
- Dosil, M., and Bustelo, X.R. (2004). Functional Characterization of Pwp2, a WD Family Protein Essential for the Assembly of the 90 S Pre-ribosomal Particle. *J. Biol. Chem.* 279, 37385–37397.
- Dragon, F., Gallagher, J.E.G., Compagnone-Post, P. a, Mitchell, B.M., Porwancher, K. a, Wehner, K. a, Wormsley, S., Settlage, R.E., Shabanowitz, J., Osheim, Y., et al. (2002). A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* 417, 967–970.
- Dubochet, J., Lepault, J., Freeman, R., Berriman, J.A., and Homo, J.-C. (1982). Electron microscopy of frozen water and aqueous solutions. *J. Microsc.* 128, 219–237.
- Dubochet, J., Adrian, M., Lepault, J., and McDowall, a (1985). Emerging techniques: Cryo-electron microscopy of vitrified biological specimens. *Trends Biochem. Sci.* 10, 143–146.
- Dubochet, J., Adrian, M., Chang, J.J., Homo, J.C., Lepault, J., McDowall, a W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* 21, 129–228.
- Dutca, L.M., Gallagher, J.E.G., and Baserga, S.J. (2011). The initial U3 snoRNA:pre-rRNA base pairing interaction required for pre-18S rRNA folding revealed by in vivo chemical probing. *Nucleic Acids Res.* 39, 5164–5180.
- Elmlund, H., Elmlund, D., and Bengio, S. (2013). PRIME: probabilistic initial 3D model generation for single-particle cryo-electron microscopy. *Structure* 21, 1299–1306.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* 66, 486–501.
- Frank, J. (2006a). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies* (Oxford University Press).
- Frank, J. (2006b). *Olfactory objects*.
- Frank, J. (2016). Generalized single-particle cryo-EM – a historical perspective. *Microscopy* 65, 3–8.
- Frank, J., Shimkin, B., and Dowse, H. (1981). SPIDER-A modular software system for electron image processing. *Ultramicroscopy* 6, 343–357.

French, S.L., Osheim, Y.N., Cioci, F., Nomura, M., and Beyer, A.L. (2003). In exponentially growing *Saccharomyces cerevisiae* cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. *Mol. Cell. Biol.* 23, 1558–1568.

Gallagher, J.E.G., Dunbar, D.A., Granneman, S., Mitchell, B.M., Osheim, Y., Beyer, A.L., and Baserga, S.J. (2004). RNA polymerase I transcription and pre-rRNA processing are linked by specific SSU processome components. *Genes Dev.* 18, 2507–2517.

Gamow, G. (1954). Possible Relation between Deoxyribonucleic Acid and Protein Structures. *Nature* 173, 318–318.

Gérczei, T., Shah, B.N., Manzo, A.J., Walter, N.G., and Correll, C.C. (2009). RNA chaperones stimulate formation and yield of the U3 snoRNA-Pre-rRNA duplexes needed for eukaryotic ribosome biogenesis. *J. Mol. Biol.* 390, 991–1006.

Grandi, P., Rybin, V., Bassler, J., Petfalski, E., Strauss, D., Marzioch, M., Schaefer, T., Kuster, B., Tschochner, H., Tollervey, D., et al. (2002). 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. *Mol. Cell* 10, 105–115.

Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *TL - 106. Proc. Natl. Acad. Sci. U. S. A.* 106 VV -, 9613–9618.

Granneman, S., Petfalski, E., Swiatkowska, A., and Tollervey, D. (2010). Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. *EMBO J.* 29, 2026–2036.

Hauk, G., and Berger, J.M. (2016). The role of ATP-dependent machines in regulating genome topology. *Curr. Opin. Struct. Biol.* 36, 85–96.

van Heel, M., Harauz, G., Orlova, E. V, Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. *J. Struct. Biol.* 116, 17–24.

Held, W.A., Mizushima, S., and Nomura, M. (1973). Reconstitution of *Escherichia coli* 30 S ribosomal subunits from purified molecular components. *J. Biol. Chem.* 248, 5720–5730.

Henras, A.K., Plisson-Chastang, C., O'Donohue, M.F., Chakraborty, A., and Gleizes, P.E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip. Rev. RNA* 6, 225–242.

Hoagland, M.B., Stephenson, M.L., Scott, J.F., Hecht, L.I., and Zamecnik, P.C. (1958). A soluble ribonucleic acid intermediate in protein synthesis. *J. Biol. Chem.* 231, 241–257.

Hohn, M., Tang, G., Goodyear, G., Baldwin, P.R., Huang, Z., Penczek, P. a, Yang, C., Glaeser, R.M., Adams, P.D., and Ludtke, S.J. (2007). SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* 157, 47–55.

Hope, H. (1988). Cryocrystallography of biological macromolecules: a generally applicable method. *Acta Crystallogr. B.* 44 (Pt 1), 22–26.

Hope, H., Frolow, F., von Böhlen, K., Makowski, I., Kratky, C., Halfon, Y., Danz, H., Webster, P., Bartels, K.S., and Wittmann, H.G. (1989). Cryocrystallography of ribosomal particles. *Acta Crystallogr. B.* 45 (Pt 2), 190–199.

- Horn, D.M., Mason, S.L., and Karbstein, K. (2011). Rcl1 protein, a novel nuclease for 18 S ribosomal RNA production. *J. Biol. Chem.* 286, 34082–34087.
- Hughes, J.M., and Ares, M. (1991). Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J.* 10, 4231–4239.
- Jakob, S., Ohmayer, U., Neueder, A., Hierlmeier, T., Perez-Fernandez, J., Hochmuth, E., Deutzmann, R., Griesenbeck, J., Tschochner, H., and Milkereit, P. (2012). Interrelationships between yeast ribosomal protein assembly events and transient ribosome biogenesis factors interactions in early pre-ribosomes. *PLoS One* 7.
- Karbstein, K. (2011). Inside the 40S ribosome assembly machinery. *Curr. Opin. Chem. Biol.* 15, 657–663.
- Karbstein, K., and Doudna, J.A. (2006). GTP-dependent formation of a ribonucleoprotein subcomplex required for ribosome biogenesis. *J. Mol. Biol.* 356, 432–443.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858.
- Kellner, N., Schwarz, J., Sturm, M., Fernandez-Martinez, J., Griesel, S., Zhang, W., Chait, B.T., Rout, M.P., Kück, U., and Hurt, E. (2016). Developing genetic tools to exploit *Chaetomium thermophilum* for biochemical analyses of eukaryotic macromolecular assemblies. *Sci. Rep.* 6, 20937.
- Kos, M., and Tollervey, D. (2010). Yeast Pre-rRNA Processing and Modification Occur Cotranscriptionally. *Mol. Cell* 37, 809–820.
- Kressler, D., Hurt, E., and Bassler, J. (2010). Driving ribosome assembly. *Biochim. Biophys. Acta - Mol. Cell Res.* 1803, 673–683.
- Krogan, N.J., Peng, W.-T.T., Cagney, G., Robinson, M.D., Haw, R., Zhong, G., Guo, X., Zhang, X., Canadien, V., Richards, D.P., et al. (2004). High-Definition Macromolecular Composition of Yeast RNA-Processing Complexes. *Mol. Cell* 13, 225–239.
- Kucukelbir, A., Sigworth, F.J., and Tagare, H.D. (2014). Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* 11, 63–65.
- Kudla, G., Granneman, S., Hahn, D., Beggs, J.D., and Tollervey, D. (2011). Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 108 VN -, 10010–10015.
- Kühlbrandt, W., and Unwin, P.N.T. (1980). Structural Analysis of Stained and Unstained Two-Dimensional Ribosome Crystals. pp. 108–116.
- Kuhn, C.D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. (2007). Functional architecture of RNA polymerase I. *Cell* 131, 1260–1272.
- Lake, J.A. (1976). Ribosome structure determined by electron microscopy of *Escherichia coli* small subunits, large subunits and monomeric ribosomes. *J. Mol. Biol.* 105, 131–159.
- Lamanna, A.C., and Karbstein, K. (2011). An RNA conformational switch regulates pre-18S rRNA cleavage. *J. Mol. Biol.* 405, 3–17.
- Lavelle, C. (2014). Pack, unpack, bend, twist, pull, push: The physical side of gene expression.

Curr. Opin. Genet. Dev. 25, 74–84.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* 10, 584–590.

Lin, J., Lai, S., Jia, R., Xu, A., Zhang, L., Lu, J., and Ye, K. (2011). Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature* 469, 559–563.

Lin, J., Lu, J., Feng, Y., Sun, M., and Ye, K. (2013). An RNA-Binding Complex Involved in Ribosome Biogenesis Contains a Protein with Homology to tRNA CCA-Adding Enzyme. *PLoS Biol.* 11.

Lu, J., Sun, M., and Ye, K. (2013). Structural and functional analysis of Utp23, a yeast ribosome synthesis factor with degenerate PIN domain. *Rna* 19, 1815–1824.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* 128, 82–97.

Madru, C., Lebaron, S., Bland, M., Delbos, L., Pipoli, J., Pasmant, E., Rety, S., and Leulliot, N. (2015). Chaperoning 5S RNA Assembly. *Genes Dev.* 29, 1432.

McKnight, S., Beyer, A., and Gall, J. (2012). Retrospective. Oscar Miller (1925-2012). *Science* 335, 1457.

Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., and Yusupov, M. (2012). One core, two shells: bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.* 19, 560–567.

Miller, O.L., and Beatty, B.R. (1969). Visualization of nucleolar genes. *Science* 164, 955–957.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* 142, 334–347.

Mizushima, S., and Nomura, M. (1970). Assembly mapping of 30S ribosomal proteins from *E. coli*. *Nature* 226, 1214–1218.

Moody, M.F. (2011). *Structural biology using electrons and X-rays an introduction for biologists* (Academic Press/Elsevier).

Mougey, E.B., O'Reilly, M., Osheim, Y., Miller, O.L., Beyer, A., and Sollner-Webb, B. (1993). The terminal balls characteristic of eukaryotic rRNA transcription units in chromatin spreads are rRNA processing complexes. *Genes Dev.* 7, 1609–1619.

Mulder, A.M., Yoshioka, C., Beck, A.H., Bunner, A.E., Milligan, R.A., Potter, C.S., Carragher, B., and Williamson, J.R. (2010). Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit. *Science* 330, 673–677.

Nierhaus, K.H., and Dohme, F. (1974). Total reconstitution of functionally active 50S ribosomal subunits from *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 71, 4713–4717.

Nirenberg, M.W., and Matthaei, J.H.H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* 47, 1588–1602.

Nogi, Y., Yano, R., and Nomura, M. (1991). Synthesis of large rRNAs by RNA polymerase II in mutants of *Saccharomyces cerevisiae* defective in RNA polymerase I. *Pnas* 88, 3962–3966.

- Nomura, M. (1973). Assembly of Bacterial Ribosomes. *Science* (80-). 179, 864–873.
- Ohtsuka, E., Moon, M.W., and Khorana, H.G. (1965). Studies on Polynucleotides. XLIII. 1 The Synthesis of Deoxyribopolynucleotides Containing Repeating Dinucleotide Sequences 2,3. *J. Am. Chem. Soc.* 87, 2956–2970.
- Osheim, Y.N., French, S.L., Keck, K.M., Champion, E.A., Spasov, K., Dragon, F., Baserga, S.J., and Beyer, A.L. (2004). Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in *Saccharomyces cerevisiae*. *Mol. Cell* 16, 943–954.
- Palade, G.E. (1955). A small particulate component of the cytoplasm. *J. Biophys. Biochem. Cytol.* 1, 59–68.
- Penczek, P.A., Grassucci, R.A., and Frank, J. (1994). The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* 53, 251–270.
- Pérez-Fernández, J., Román, A., De Las Rivas, J., Bustelo, X.R., and Dosil, M. (2007). The 90S preribosome is a multimodular structure that is assembled through a hierarchical mechanism. *Mol. Cell. Biol.* 27, 5414–5429.
- Petes, T.D. (1979). Yeast ribosomal DNA genes are located on chromosome XII. *Proc. Natl. Acad. Sci. U. S. A.* 76, 410–414.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Phipps, K.R., Charette, J.M., and Baserga, S.J. (2011). The small subunit processome in ribosome biogenesis-progress and prospects. *Wiley Interdiscip. Rev. RNA* 2, 1–21.
- Pintilie, G., Zhang, J., Chiu, W., and Gossard, D. (2009). Identifying components in 3D density maps of protein nanomachines by multi-scale segmentation. 2009 IEEE/NIH Life Sci. Syst. Appl. Work. LiSSA 2009 44–47.
- Rabl, J., Leibundgut, M., Ataide, S., Haag, A., and Ban, N. (2011). Crystal Structure of the Eukaryotic 40S Ribosomal Subunit in Complex with Initiation Factor 1. *Science* (80-). 331, 730.
- Russel, D., Lasker, K., Webb, B., Velazquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B., and Sali, A. (2012). Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* 10, 1–5.
- Sardana, R., Liu, X., Granneman, S., Zhu, J., Gill, M., Papoulas, O., Marcotte, E.M., Tollervey, D., Correll, C.C., and Johnson, A.W. (2015). The DEAH-box Helicase Dhr1 Dissociates U3 from the Pre-rRNA to Promote Formation of the Central Pseudoknot. *PLoS Biol.* 13, 1–25.
- Sashital, D.G., Greeman, C.A., Lyumkis, D., Potter, C.S., Carragher, B., and Williamson, J.R. (2014). A combined quantitative mass spectrometry and electron microscopy analysis of ribosomal 30S subunit assembly in *E. coli*. *Elife* 3, e04491.
- Scheres, S.H.W. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* 180, 519–530.
- Schlueder, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., et al. (2000). Structure of Functionally Activated Small

Ribosomal Subunit at 3.3 Å Resolution. *Cell* 102, 615–623.

Shaikh, T.R., Gao, H., Baxter, W.T., Asturias, F.J., Boisset, N., Leith, A., and Frank, J. (2008a). SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat. Protoc.* 3, 1941–1974.

Shaikh, T.R., Trujillo, R., LeBarron, J.S., Baxter, W.T., and Frank, J. (2008b). Particle-verification for single-particle, reference-based reconstruction using multivariate data analysis and classification. *J. Struct. Biol.* 164, 41–48.

Simonović, M., and Steitz, T.A. (2009). A structural view on the mechanism of the ribosome-catalyzed peptide bond formation. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1789, 612–623.

Sorzano, C.O.S., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H.W., Carazo, J.M., and Pascual-Montano, A. (2004). XMIPP: A new generation of an open-source image processing package for electron microscopy. *J. Struct. Biol.* 148, 194–204.

Soudet, J., Gélugne, J.-P., Belhabich-Baumais, K., Caizergues-Ferrer, M., and Mougin, A. (2010). Immature small ribosomal subunits can engage in translation initiation in *Saccharomyces cerevisiae*. *EMBO J.* 29, 80–92.

Stokes, J.M., Davis, J.H., Mangat, C.S., Williamson, J.R., and Brown, E.D. (2014). Discovery of a small molecule that inhibits bacterial ribosome biogenesis. *Elife* 3, e03574.

Strunk, B.S., Loucks, C.R., Su, M., Vashisth, H., Schilling, J., Iii, C.L.B., and Karbstein, K. (2012). Initiation by 40S Assembly Intermediates. 333, 1449–1453.

Taddei, G. (1972). RIBOSOME ARRANGEMENT DURING SICULA O ~~~ NES ~ S OF. 70, 285–292.

Tan, R.Z., Oudenaarden, A. Van, and van Oudenaarden, A. (2010). Transcript counting in single cells reveals dynamics of rDNA transcription. *Mol. Syst. Biol.* 6, 358.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* 157, 38–46.

Thomas, S., Keller, C., Szyk, A., Cannon, J., and Laronde-Leblanc, N.A. (2011). Structural insight into the functional mechanism of Nep1/Emg1 N1-specific pseudouridine methyltransferase in ribosome biogenesis. *Nucleic Acids Res.* 39, 2445–2457.

Tishchenko, S., Nikonova, E., Nikulin, A., Nevskaya, N., Volchkov, S., Piendl, W., Garber, M., and Nikonov, S. (2006). Structure of the ribosomal protein L1-mRNA complex at 2.1 Å resolution: common features of crystal packing of L1-RNA complexes. *ACTA CRYSTALLOGR., SECT. D* 62, 1545–1554.

Trapman, J., Retel, J., and Planta, R.J. (1975). Ribosomal precursor particles from yeast. *Exp. Cell Res.* 90, 95–104.

Traub, P., and Nomura, M. (1968). Structure and function of *E. coli* ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. *Proc. Natl. Acad. Sci. U. S. A.* 59, 777–784.

Unwin, P.N.T. (1977). Three-dimensional model of membrane-bound ribosomes obtained by electron microscopy. *Nature* 268, 495–500.

Warner, J.R. (1999). The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.*

24, 437–440.

Wasilewski, S., and Rosenthal, P.B. (2014). Web server for tilt-pair validation of single particle maps from electron cryomicroscopy. *J. Struct. Biol.* 186, 122–131.

Williamson, J.R. (2009). The Ribosome at Atomic Resolution. *Cell* 139, 1041–1043.

Wilson, D.N. (2014). Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev. Microbiol.* 12, 35–48.

Wimberly, B.T., Brodersen, D.E., Clemons Jr, W.M., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* 407, 327–339.

Woolford, J. (2015). Assembly of ribosomes in eukaryotes. *RNA* 21, 766–768.

Woolford, J.L., and Baserga, S.J. (2013). Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics* 195, 643–681.

Yang, B., Wu, Y.-J., Zhu, M., Fan, S.-B., Lin, J., Zhang, K., Li, S., Chi, H., Li, Y.-X., Chen, H.-F., et al. (2012a). Identification of cross-linked peptides from complex samples. *Nat. Methods* 9, 904–906.

Yang, Z., Fang, J., Chittuluru, J., Asturias, F.J., and Penczek, P.A. (2012b). Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* 20, 237–247.

Yonath, A.E. (2009). Hibernating Bears , Antibiotics and the Evolving Ribosome. *Nobel Lect.* 211–237.

Yoshioka, C., Lyumkis, D., Carragher, B., and Potter, C.S. (2013). Maskitron: Interactive, web-based classification of single-particle electron microscopy images. *J. Struct. Biol.* 182, 155–163.

Zamecnik, P.C., and Keller, E.B. (1954). Relation between phosphate energy donors and incorporation of labeled amino acids into proteins. *J. Biol. Chem.* 209, 337–354.

Zhang, C., Lin, J., Liu, W., Chen, X., Chen, R., and Ye, K. (2014). Structure of Utp21 Tandem WD Domain Provides Insight into the Organization of the UTPB Complex Involved in Ribosome Synthesis. *PLoS One* 9, e86540.

Zhang, L., Lin, J., and Ye, K. (2013). Structural and functional analysis of the U3 snoRNA binding protein Rrp9. *Rna* 19, 701–711.

Zheng, S., Lan, P., Liu, X., and Ye, K. (2014). Interaction between ribosome assembly factors Krr1 and Faf1 is essential for formation of small ribosomal subunit in yeast. *J.Biol.Chem.* 289, 22692–22703.

Curriculum Vitae

Martin Turk

Gene Center, Ludwig-Maximilians-Universität München

Feodor-Lynen.Str. 25, Munich, Germany

Phone: +49 176 996 32 696 Email: martin.turk@gmail.com

PERSONAL

Name: Martin Turk
Date of birth: 20.4.1985
Place of birth: Ljubljana, Slovenia
Nationality: Slovenian

EDUCATION

Biochemistry at the University of Ljubljana, Slovenia 2003 – 2009
• Diploma: "Activity of protein kinase A in isolated rat astrocyte monitored by a fluorescence nanosensor" June 2009

PUBLICATIONS

Kornprobst M¹, Turk M¹, Kellner N, Cheng J, Flemming D, Kos-Braun IC, Kos M, Thoms M, Berninghausen O, Beckmann R*, Hurt E*.
"Architecture of the 90S pre-ribosome: a structural view on the birth of the eukaryotic ribosome"
Cell, accepted 29.4.2016

van Dam TJ, Townsend MJ, Turk M, Schlessinger A, Sali A, Field MC, Huynen MA.
"Evolution of modular intraflagellar transport from a coatomer-like progenitor."
Proc Natl Acad Sci USA 110(17), 6943-6948, 2013

CURRENT OCCUPATION

IMPRS-LS Ph.D. Student at the Beckmann Lab Sep 2010 – present
• Project 1: Structural Characterization of the RSC chromatin remodeler 2010 – Feb 2014
• Project 2: 90S pre-ribosome Feb 2014 – present

VISITING STUDENT

Evolutionary origins of eukaryotic intracellular trafficking systems Aug 2009 – Feb 2010
At Andrej Šali's Lab, UCSF

PRAXIS

Side-chain optimization by considering hydrogen bond formation Mar – Aug 2010
At Dušan Turk's Lab, IJS

Summer project: expression of unnatural amino acids in *E.coli* Aug – Sep 2005
At Nediljko Budiša's Lab, MPI-B

WORKSHOPS

NRAMM Nov 9 – 14, 2014

Workshop on Advanced Topics in EM Structure Determination
Scripps, San Diego

EMBO Practical Course
Image Processing for cryo-EM
Birkbeck, London
Selected for poster talk on the RSC remodeler

Sep 3 – 13, 2013

EMBO Practical Course
Structural Characterization of Macromolecular Complexes
EPN, Grenoble

June 4 – 9, 2012

SCHOLARSHIPS

Zois Scholarship, merit-based state scholarship of Slovenia

1999 – 2008

SKILLS

Biochemistry

- Large scale endogenous purification from yeast
- Purification and analysis of recombinant proteins from *E. Coli*
- Nucleosome reconstitution

EM

- Negative staining
- Basic usage of LVEM5 and FEI Morgagni microscopes
- 2D classification, *ab initio* models
- Software: Sparx, Eman2, Relion, Spider, Simple PRIME, Imagic, Frealign, Xmipp

Computer literacy

- Python (with numpy, scipy), bash and zshell
- Reading level: c, fortran, perl
- Wrote numerous conversion scripts mainly in python for switching between different software packages, selecting from .star files, writing/reading to/from .hdf, .mrc headers
- Wrote encloseVolume command for UCSF Chimera

SCIENTIFIC INTERESTS

- Finding out how things work
- Problem solving
- Making tools

REFERENCES

Prof. Roland Beckmann
Gene Center, LMU
Feodor-Lynen.Str. 25
81377 München
+49 89 2180 76900
beckmann@genzentrum.lmu.de

Prof. Mario Halič
Gene Center, LMU
Feodor-Lynen.Str. 25
81377 München
+49 89 2180 76962
halič@genzentrum.lmu.de

Prof. Friedrich Förster
MPI-B
Am Klopferspitz 18
82152 Martinsried
+49 89 8578 2632
foerster@biochem.mpg.de
and
Bijvoet Center for
Biomolecular Research
Utrecht University
Padualaan 8
3584 CH Utrecht
The Netherlands

Prof. Andrej Šali
UCSF
1700 4th Street
94143 San Francisco
+1 415 514 4227
sali@salilab.org

Acknowledgments

First of all I'd like to thank everyone who contributed to the successful outcome of my thesis. It was hard work and the support was much appreciated. I don't know how to thank enough my girlfriend Ana, who has patiently suffered with me throughout the full 6 years of my PhD. I hope she has learned a thing or two in the process of being forced to listen to my many monologues on how something works whether it was related to more technical aspects or to the inner workings of a cell. Through explaining things as simply as possible I have learned a lot in return and gained a much better understanding about many topics. I am extremely grateful to have such a great listener.

I'd like to thank Professor Beckmann for giving me the complete freedom in handling both my PhD projects. It often was not easy, but in the process of discovering things on my own I've learned a lot and this knowledge will stay with me for which I am very grateful. This kind of knowledge is different in its nature to the knowledge received on a silver platter that quickly leads to successful experiments. Unfortunately the first project on the structural characterization of the RSC chromatin remodeler did not work out, it was an interesting project, but it did bring me closer to Professor Halič who worked on related topics and who adopted me for a short while in his lab. This gave me the opportunity to look at things from a different perspective and I learned a lot on properly constructing controls for my experiments in his environment.

I am again grateful to Professor Beckmann to have given me a second opportunity at another difficult project the 90S pre-ribosome complex after the first project had failed, allowing me to show my so far learned skills and this time successfully apply them. Our conversations often not related to the project at hand have considerably broadened my horizon and perhaps most importantly taught me not to jump too quickly to conclusions. Often I was right, but sometimes it turned out I was less so, and this is a valuable lesson for the future.

I have given first row tickets to my parents, my brother and both my grandparents to what would be my PhD show, then dragged them into the theatre, on more than one occasion, and made them observe how the drama unfolds. I am afraid often they were left little choice. Six years is a long time and I am happy everyone put up with me that long. There were also many pleasant moments and I'd like to thank Professor Wendler for the many interesting conversations we had and I hope to be able to make her draw a smile when she is reading this, I know smiling is not easy. I have extremely enjoyed interesting conversation with my grandmother, who taught me much about life and who has never stopped sending positive thoughts my way. My grandfather who is tending towards holding long monologs himself, he is by far not an easy listener, but in time I've learned to trick him, yes he can listen if he wants. I'd like to thank my parents who stood by my side even after I have left the house and for keeping my room my own. When two people live 5 years in a one room apartment one certainly appreciates coming back to his own bedroom during visits, which is the size of his kitchen and living room combined. I think in time our relationship gained in depth, but this is only natural. And my younger brother who I have not spent as much time with as I'd wished for, but things change. I think he will too soon realize what making a living in life means.

Before this becomes another one of my monologues I have to stop myself and finish by thanking everyone for reading this far. A final thought I wish to share and then I'll stop. It is impossible to know everything, but one can make an effort to stay less ignorant and this is very important. I do hope some people will find themselves in this thought. Thanks again.

Munich, June 15th 2016.